İstanbul Gelişim Üniversitesi
Sağlık Bilimleri Dergisi
(IGUSABDER)

Sayı / Issue: 29

Sağlık Uygulamalarında Yapay Zekâ Özel Sayısı
Special Issue on Artificial Intelligence in Health Applications

Yıl / Year: 2026

Istanbul Gelisim University
Journal of Health Sciences
(IGUSABDER)

ISSN: 2536-4499
e-ISSN: 2602-2605

İstanbul Gelişim Üniversitesi kurumsal yayını olan Sağlık Bilimleri Dergisi, yılda üç kez yayımlanan uluslararası hakemli bir dergidir. Makalelerdeki görüş, düşünce, varsayım veya öneriler eser sahiplerine aittir; İstanbul Gelişim Üniversitesi sorumlu tutulamaz.

The Journal of Health Sciences is an international peer–reviewed journal and will be published three times a year. The opinions, thoughts, postulations or proposals within the articles are but reflections of the authors and do not, in any way, represent those of the Istanbul Gelisim University.

İLETİŞİM BİLGİLERİ / COMMUNICATION:

İstanbul Gelişim Üniversitesi, Sağlık Bilimleri Fakültesi,
Cihangir Mah. Şehit Jandarma Komando Er Hakan Öner Sk.
No: 1, 34310 Avcılar / İstanbul / TÜRKİYE

Tel: +90 212 4227000 **Dahili** 7333
Belgeç: +90 212 4227401
E-posta: igusabder@gelisim.edu.tr
Ağ sayfası: https://igusabder.gelisim.edu.tr
https://twitter.com/igusabder
Twitter: @igusabder

Baskı ve cilt:
Printing and binding:

Servet İşler
Sertifika No. 40352
Tel: +90 212 5939467
E-posta: islercopy@hotmail.com

İSTANBUL GELİŞİM ÜNİVERSİTESİ SAĞLIK BİLİMLERİ DERGİSİ: İNDEKSLENME VE HARMANLANMA BİLGİLERİ /
ISTANBUL GELISIM UNIVERSITY JOURNAL OF HEALTH SCIENCES: INFORMATION ABOUT INDEXING AND HARVESTING

DİZİNLENME / INDEXING

Dergimiz, TÜBİTAK ULAKBİM TR Dizin tarafından 2020 yılından bu yana dizinlenmektedir.

Dergimiz, Türkiye Atıf Dizini tarafından 2017 yılından bu yana dizinlenmektedir.

Sağlık Bilimleri Alan İndeksi

HARMANLANMA / HARVESTING

Dergimizin içeriği Dspace@Gelişim Kurumsal Açık Erişim Arşivi (https://acikerisim.gelisim.edu.tr) ve Avrupa Komisyonu'nun OpenAIRE 2020 Açık Erişim Projesi (The OpenAIRE2020 Project) (https://www.openaire.eu/openaire2020-project) tarafından harmanlanmaktadır.

İSTANBUL GELİŞİM ÜNİVERSİTESİ SAĞLIK BİLİMLERİ DERGİSİ (IGUSABDER)
Uluslararası Hakemli Dergi
Sayı 29 • Temmuz • 2026 • ISSN: 2536-4499 • e-ISSN: 2602-2605
Sağlık Uygulamalarında Yapay Zekâ Özel Sayısı

İLETİŞİM:
İstanbul Gelişim Üniversitesi, Sağlık Bilimleri Fakültesi,
Cihangir Mah. Şehit Jandarma Komando Er Hakan Öner Sk. No: 1,
34310 Avcılar / İstanbul / TÜRKİYE
Tel: +90 212 4227000 **Dahili:** 7333
Belgeç: +90 212 4227401
E-posta: igusabder@gelisim.edu.tr
Ağ sayfası: https://igusabder.gelisim.edu.tr
Twitter: @igusabder

ISTANBUL GELISIM UNIVERSITY JOURNAL OF HEALTH SCIENCES (IGUSABDER)
International Peer–Reviewed Journal
Issue 29 • July • 2026 • ISSN: 2536-4499 • e-ISSN: 2602-2605
Special Issue on Artificial Intelligence in Health Applications

COMMUNICATION:
Istanbul Gelisim University, Faculty of Health Sciences,
Sehit Jandarma Komando Er Hakan Oner Sk. No: 1,
34310 Avcilar / Istanbul / TÜRKİYE
Phone: +90 212 4227000 **Ext.** 7333
Fax: +90 212 4227401
E-mail: igusabder@gelisim.edu.tr
Web page: https://igusabder.gelisim.edu.tr
Twitter: @igusabder

BİLİMSEL DANIŞMA KURULU

BU SAYININ HAKEMLERİ

SCIENTIFIC ADVISORY BOARD

REFEREES FOR THIS ISSUE

Editörden

Değerli Okurlar,

İstanbul Gelişim Üniversitesi Sağlık Bilimleri Dergisi'nin ***"Sağlık Uygulamalarında Yapay Zekâ"*** başlıklı özel sayısını sizlerle buluşturmaktan büyük mutluluk duyuyoruz.

Dergimiz, 2017 yılında başladığı yayın hayatını bilimsel nitelik, etik yayıncılık ve disiplinlerarası yaklaşım temelinde sürdürmektedir. Kuruluşundan bu yana sağlık bilimlerinin farklı alanlarında üretilen güncel ve özgün bilgiyi bilim dünyasıyla buluşturan dergimiz, Dr. Öğr. Üyesi A. Yüksel BARUT'un editörlüğünde akademik yayıncılık alanındaki gelişimini, ulusal ve uluslararası görünürlüğünü güçlendirmeye devam etmektedir. Bu özel sayı ise dergimizin dokuz yıllık bilimsel yayıncılık birikimini günümüzün en hızlı gelişen alanlarından biri olan yapay zekâ ile bir araya getirmektedir.

Yapay zekâ; hastalıkların erken tanılanması, klinik karar süreçlerinin desteklenmesi, sağlık verilerinin değerlendirilmesi, bireyselleştirilmiş uygulamaların geliştirilmesi ve sağlık hizmetlerinin daha etkin biçimde sunulması açısından önemli fırsatlar sağlamaktadır. Bununla birlikte veri güvenliği, mahremiyet, algoritmik yanlılık, açıklanabilirlik, etik sorumluluk ve insan denetimi gibi konular da bu dönüşümün ayrılmaz bir parçasıdır. Bu nedenle yapay zekânın sağlık alanına entegrasyonu, insan merkezli, etik, güvenilir ve kanıta dayalı bir yaklaşımla ele alınmalıdır.

Bu özel sayıda, yapay zekânın sağlık bilimlerindeki kullanım alanlarını, sunduğu fırsatları ve beraberinde getirdiği sınırlılıkları farklı disiplinlerin bakış açılarıyla değerlendiren çalışmalara yer verilmiştir. Çalışmaların araştırmacılar, sağlık profesyonelleri, eğitimciler ve öğrenciler için güncel bir kaynak oluşturmasını, yeni araştırmalara ve disiplinlerarası işbirliklerine zemin hazırlamasını diliyoruz. Üniversitemizin "Araştırma Üniversitesi" olma vizyonuyla da örtüşen bu özel sayının, sağlık bilimlerinde yenilikçi ve disiplinlerarası araştırmaların geliştirilmesini teşvik edeceğine inanıyoruz.

Özel sayının hazırlanması sürecindeki destekleri ve değerli katkıları için dergi editörümüz Dr. Öğr. Üyesi A. Yüksel BARUT'a, diğer editör yardımcımız Prof. Dr. S. Arda ÖZTÜRKCAN'a; bilimsel değerlendirmeleriyle sayının niteliğine katkı sunan hakemlerimize, tüm dergi ekibimize, editör kurulumuza ve çalışmalarını bizlerle paylaşan tüm yazarlarımıza teşekkür ederim. Bu özel sayının, sağlık uygulamalarında yapay zekânın bilimsel, etik ve sorumlu kullanımına katkı sağlamasını, yeni araştırmalara ve disiplinlerarası işbirliklerine zemin hazırlamasını temenni eder, keyifli okumalar dilerim.

Doç. Dr. Hatice Merve BAYRAM

Özel Sayı Misafir Editörü

From the Editor

Dear Readers,

We are delighted to present to you the special issue of the Istanbul Gelisim University Journal of Health Sciences entitled ***"Artificial Intelligence in Health Applications"***.

Since its launch in 2017, our journal has continued its publication journey on the basis of scientific quality, ethical publishing, and an interdisciplinary approach. Since its establishment, our journal has brought current and original knowledge produced in various fields of health sciences to the scientific community and, under the editorship of Asst. Prof. A. Yüksel BARUT, continues to advance in academic publishing and strengthen its national and international visibility. This special issue brings together our journal's nine years of scientific publishing experience with artificial intelligence, one of today's most rapidly developing fields.

Artificial intelligence offers significant opportunities for the early diagnosis of diseases, support of clinical decision-making processes, evaluation of health data, development of personalized practices, and more effective delivery of healthcare services. At the same time, issues such as data security, privacy, algorithmic bias, explainability, ethical responsibility, and human oversight are also integral to this transformation. Therefore, the integration of artificial intelligence into healthcare should be addressed through a human-centered, ethical, reliable, and evidence-based approach.

This special issue features studies that evaluate, from the perspectives of different disciplines, the applications of artificial intelligence in health sciences, the opportunities it offers, and the limitations it entails. We hope that these studies will serve as an up-to-date resource for researchers, healthcare professionals, educators, and students and will lay the groundwork for new research and interdisciplinary collaborations. We believe that this special issue, which also aligns with our university's vision of becoming a "Research University," will encourage the development of innovative and interdisciplinary research in health sciences.

I would like to thank our journal editor, Asst. Prof. A. Yüksel BARUT, and our fellow assistant editor, Prof. S. Arda ÖZTÜRKCAN, for their support and valuable contributions during the preparation of this special issue; our reviewers, whose scientific evaluations contributed to the quality of the issue; the entire journal team; our editorial board; and all the authors who shared their work with us. I hope that this special issue will contribute to the scientific, ethical, and responsible use of artificial intelligence in health practices and lay the groundwork for new research and interdisciplinary collaborations. I wish you an enjoyable read.

Assoc. Prof. Hatice Merve BAYRAM

Guest Editor

İstanbul Gelişim Üniversitesi, Sağlık Bilimleri Fakültesi'nin aşağıdaki Bölümleri, Almanya merkezli **Accreditation Agency in Health and Social Sciences / Akkreditierungsagentur im Bereich Gesundheit und Soziales (AHPGS)** tarafından

Beslenme ve Diyetetik (TR-İNG),
Fizyoterapi ve Rehabilitasyon (TR-İNG),
Hemşirelik (TR-İNG) bölümleri 2030 yılına kadar;

Çocuk Gelişimi (TR-İNG),
Ergoterapi,
Odyoloji,
Sağlık Yönetimi,
Sosyal Hizmet (TR-İNG) bölümleri 2018-2023 yılları arasında koşulsuz olarak akredite edilmiştir.

&

The Following Departments of Istanbul Gelisim University, Faculty of Health Sciences have been unconditionally accredited by the Germany based ***Accreditation Agency in Health and Social Sciences / Akkreditierungsagentur im Bereich Gesundheit und Soziales (AHPGS)***

Nutrition and Dietetics (Turkish-English Tracks)
Physical Therapy and Rehabilitation (Turkish-English Tracks),
Nursing (Turkish-English Tracks) *departments until 2030;*

Audiology,
Child Development (Turkish-English Tracks),
Health Management,
Occupational Theraphy,
Social Service (Turkish-English Tracks) *departments were accredited unconditionally between 2018-2023.*

İÇİNDEKİLER / CONTENTS

Boric Acid-Induced Apoptosis in Breast Cancer Subtypes: Comparative Gene Expression and Machine Learning Study

Şerife Buket BOZKURT POLAT[*], Andaç Batur ÇOLAK[**], Seyyid Mehmet BULUT[***]

Abstract

Aim: Breast cancer is a diverse illness that arises from the interplay of outside factors and genetic predisposition. Recently, artificial intelligence approaches have enabled the systematic analysis of potential anticancer agents such as boric acid by modeling the functions of apoptotic regulators. This study proposes an artificial intelligence-based analytical approach to evaluate how boric acid (B) affects cell survival and cell death through apoptosis pathway modulation.

Method: Viability of cells was determined by applying boric acid at various concentrations of 0.1, 1, 10, 50, 100, and 1000 ng/mL to cells at 24 h and 72 h using the cell viability test. Lastly, total RNA was isolated at 24 and 72 hours, and messenger RNA (mRNA) expression levels of specific genes were established by quantitative PCR (qPCR). An analysis was carried out using artificial intelligence based upon the Extreme Gradient Boosting (XGBoost) solution, a technique that enables the development of a pair of predictive models for cell viability and relative mRNA expression.

Results: After 24- and 72-hour treatments, boric acid reduced cell viability in MCF-7 and MDA-MB-231 cells in a rate- and time-sensitive manner in relation to the unaffected category. The triple-negative breast cancer phenotype was shown to be more susceptible to apoptosis at both time periods, as evidenced by the increased lethal consequences of boric acid in MDA-MB-231 cells compared to MCF-7 cells. However, boric acid was found to induce both intrinsic and extrinsic apoptosis in both breast cancer cell lines. Specifically for internal apoptosis signaling, ENDOG, BAX, Caspase-9, and Caspase-3 mRNA expression increased, while BCL-2 mRNA expression decreased, and the BAX/BCL-2 ratio was determined to shift towards apoptosis. Furthermore, apoptotic effects were detected in MCF-7 cells at 24 and 72 hours into the experiment, with this effect being particularly higher in MDA-MB-231 cells at 24 hours. The results obtained showed that boric acid induced the extrinsic apoptosis pathway by increasing mRNA FAS, FASLG, and CASP8 mRNA expression levels at 72 hours in MCF-7 cells and at 24 hours in MDA-MB-231 cells. However, boric acid increased MYC expression in MDA-MB-231 cells at 24 hours, while this effect was observed in MCF-7 cells at 72 hours of the experiment. The optimized XGBoost models were insufficient and showed highly non-homogeneous Margin of Deviation (MoD) values, with cell viability in the range of ±100% and mRNA expression spanning -4000% to +2000%.

Conclusions: The results obtained indicate that boric acid regulates cellular survival and death in breast cancer subtypes through MYC-dependent mechanisms and that this effect exhibits a more pronounced

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 16.12.2025 **Kabul / Accepted:** 20.07.2026
DOI: https://doi.org/10.38079/igusabder.1843261
[*] Assoc. Prof., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Biochemistry, Niğde, Türkiye.
E-mail: buketbozkurt@yahoo.com ORCID https://orcid.org/0000-0002-8641-2844
[**] Assoc. Prof., Niğde Ömer Halisdemir University, Mechanical Engineering Department, Niğde, Türkiye.
E-mail: andacbaturcolak@hotmail.com ORCID https://orcid.org/0000-0001-9297-8134
[***] Res. Asst., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Biochemistry, Niğde, Türkiye.
E-mail: sbulut@ohu.edu.tr ORCID https://orcid.org/0000-0001-5020-4210

antitumor effect in MDA-MB-231 cells. So, it is concluded that boric acid may be a promising therapeutic agent for breast cancer if supported by additional research. Moreover, the inability of the current AI models also accentuates the complexity and non-linearity of the cellular response and suggests that the quantitative biological outcome can be better predicted with a broader multi-feature set.

Keywords: Boric acid, MCF-7, MDA-MB-231, XGBoost.

Meme Kanseri Alt Tiplerinde Borik Asit ile İndüklenen Apoptoz: Karşılaştırmalı Gen Ekspresyonu ve Makine Öğrenmesi Çalışması

Öz

Amaç: Meme kanseri, dış faktörlerin ve genetik yatkınlığın etkileşimi sonucu ortaya çıkan çok yönlü bir hastalıktır. Son zamanlarda yapay zeka yaklaşımları, apoptotik düzenleyicilerin işlevlerini modelleyerek borik asit gibi potansiyel antikanser ajanlarının sistematik analizini mümkün kılmaktadır. Bu çalışma, borik asit (B)'in apoptoz yolak modülasyonu üzerinden hücre sağkalımı ve hücre ölümünü nasıl etkilediğini değerlendirmek amacıyla yapay zeka tabanlı analitik bir yaklaşım önermektedir.

Yöntem: Hücre canlılığı, MCF-7 ve MDA-MB-231 meme kanseri hücre hatlarında 0.1, 1, 10, 50, 100 ve 1000 ng/mL konsantrasyonlarında borik asit uygulamasından 24 ve 72 saat sonra hücre canlılığı testi ile değerlendirildi. Son olarak, 24 ve 72 saatte toplam RNA izole edildi ve belirli genlerin mRNA ekspresyon seviyeleri kantitatif PCR (qPCR) ile belirlendi. Deneysel kantifikasyon için yapay zeka tabanlı bir analitik yaklaşım kullanıldı. Sonrasında hücre canlılığı ve mRNA ekspresyonunun iki bağımsız tahmin modelini oluşturmak için Extreme Gradient Boosting (XGBoost) çerçevesi kullanıldı.

Bulgular: Deneyin 24 ve 72. saatlerinde, borik asit, kontrol grubuna kıyasla, doz ve zamana bağlı olarak MCF-7 ve MDA-MB-231 hücre canlılığında azalmaya yol açtı. Borik asidin sitotoksik etkileri, MDA-MB-231 hücrelerinde MCF-7 hücrelerine göre daha güçlü gözlemlendi, bu da her iki zaman noktasında da üçlü negatif meme kanseri fenotipinde apoptoza karşı daha yüksek bir duyarlılığa işaret etmektedir. Bununla birlikte borik asit, her iki meme kanseri hücresinin iç ve dış apoptozunu indüklediği tespit edilmiştir. Özellikle iç apoptoz sinyallemesi için, ENDOG, BAX, Caspase-9 ve Caspase-3 mRNA ekspresyonu artarken, BCL-2 mRNA ekspresyonu ise azaldı ve BAX/BCL-2 oranının apoptoza doğru yön değiştirdiği belirlendi. Ayrıca apoptotik etkiler, deneyin 24. ve 72. Saatlerinde MCF-7 hücrelerinde belirlenirken özellikle bu etkinin deneyin 24. saatinde MDA-MB-231 hücrelerinde daha yüksek olduğu ortaya çıktı. Elde edilen sonuçlar, borik asitin MCF-7 hücrelerinde 72. saatte ve MDA-MB-231 hücrelerinde ise 24. saatte mRNA FAS, FASLG ve CASP8 mRNA ekspresyon seviyelerinin artarak dış apoptoz yolunu indüklediğini gösterdi. Bununla birlikte borik asit, MDA-MB-231 hucrelerinde 24. saatte MYC ekspresyonunu artırırken bu etki MCF-7 hücrelerinde deneyin 72. saatinde belirlendi. Çalışma, optimize edilmiş XGBoost modellerinin tahmin doğruluğunun yetersiz olduğunu, hem hücre canlılığı (±%100'e kadar) hem de mRNA ifadesi (-%4000'den +%2000'e kadar değişen) için son derece tutarsız sapma marjı değerlerine sahip olduğunu ortaya koydu.

Sonuç: Elde edilen sonuçlar, borik asidin meme kanseri alt tiplerinde hücresel sağkalım ve ölümü MYC-bağımlı mekanizmalar üzerinden düzenlediğini ve bu etkinin MDA-MB-231 hücrelerinde daha belirgin antitümoral etki ortaya koyduğunu göstermektedir. Bu nedenle, daha ileri gelişmiş deneylerle desteklenen borik asit, meme kanserini tedavi etmek için kullanılabilecek potansiyel bir terapötik ajandır. Ayrıca, mevcut yapay zeka modellerinin başarısızlığı, hücresel yanıtın son derece karmaşık ve son derece doğrusal olmayan bir yapıya sahip olduğunu göstermekte, bu da nicel biyolojik sonuç tahminlerini mümkün kılan daha geniş, çok özellikli bir setin önemli bir gereklilik olduğunu ortaya koymaktadır.

Anahtar Sözcükler: Borik asit, MCF-7, MDA-MB-231, XGBoost.

Introduction

Breast cancer subtypes include hormone receptor-positive, HER2-enriched, and triple-negative breast cancer, each with distinct genetic signatures. LncRNAs like MIR200CHG promote tumors by binding regulatory proteins[1]. Treatment resistance develops through gene mutations, nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) activation, and sphingolipid signaling[2,3]. Sphingosine kinase activity correlates with chemoresistance, while Sphk2 inhibition may induce apoptosis. Cancer cells resist apoptosis through protein overproduction or mitochondrial dysfunction. NF-κB enables drug-resistant cell survival[4,5]. ADAM12 enhances death in the tumor microenvironment. Preclinical studies show promise in targeting apoptotic programs. Apoptosis induction through mitochondrial, FS-7-associated surface antigen (Fas) and Fas Ligand (FasL); FasL/Fas, phosphoinositide 3-kinase/protein kinase B (PI3K/AKT), and mitogen-activated protein kinases (MAP kinase) pathways advances therapy[6-8]. Breast cancer's heterogeneity affects treatment efficacy, while resistance stems from gene expression variations. Identifying cell survival drivers is crucial for reducing resistance. Artificial intelligence (AI) advances cancer research by analyzing multi-omics data to identify pathways where traditional analyses fail[9,10]. AI enables patient risk stratification through clinical metrics analysis, supporting personalized medicine[11]. For drug selection, AI speeds development using computer assisted engineering to find promising drugs[12,13]. AI technology with three-dimensional models improves prediction of drug effects, enhancing preclinical screening[14]. AI recognizes biomarkers, processes data, and elevates image analysis for improved diagnostics[15,16]. AI models have reduced time-to-hypothesis and integrated heterogeneous data into clinical decisions, offering precision in oncology[9]. AI in cancer research applies to molecular regulator discovery, drug efficacy evaluation, and biomedical data handling.

Boric acid is a naturally occurring, water-soluble form of boron—a trace element that is not synthesized endogenously but is obtained through dietary intake and rapidly absorbed from the gastrointestinal tract, entering the bloodstream predominantly as boric acid (98.4%) with only 1.6% as borate anion, with a recommended upper intake level of 20 mg/day in humans[17,18]. Chemically, boric acid (H_3BO_3) functions as a Lewis acid with a pKa of approximately 9.24; at physiological pH 7.4, it exists predominantly in its undissociated monomeric form rather than as borate anion[18]. At the molecular level, boric acid modulates intracellular Ca^{2+} signaling and activates the Nrf2 antioxidant pathway, thereby regulating transcription factors involved in anti-inflammatory gene induction[19,20]. Boric acid has well-documented antiseptic, anti-inflammatory, and antimicrobial properties; specifically, it inhibits lipopolysaccharide-induced TNF-α formation, suppresses pro-inflammatory cytokines, and enhances anti-inflammatory cytokine expression under conditions of inflammatory and oxidative stress[17,20]. In mineralized tissues, physiological concentrations of boric acid upregulate key osteogenic markers including RunX2, BSP, OCN, OPN, COL-I, and BMP-4/-6/-7, and induce tuftelin — a differentiation factor critical for the initial stages of bone and tooth formation and mineralization[18,19]. Literature has explored its therapeutic potential in cancer treatment, including its biocompatibility and bioavailability[21]. Studies show boric acid

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

inhibits cancer cell growth, with research demonstrating dose-dependent suppression of breast cancer cell proliferation. While calcium fructoborate induced apoptosis through caspase-3 activity, boric acid inhibited growth without clear apoptosis induction, suggesting different mechanisms requiring further study[21]. In prostate cancer cells, boric acid induced senescence-like changes, reduced cell cycle proteins, and metastatic potential, demonstrating chemopreventive properties[22]. The focus on boric acid in breast cancer treatment stems from its proven growth-inhibiting effects, low toxicity, and potential synergy with other therapies. Studies show beneficial effects when combining boric acid with agents like paclitaxel in polymeric delivery systems[23]. Given breast cancer's heterogeneity, new drugs targeting multiple cellular pathways are needed. While boric acid shows promise, further research is needed to understand its role in breast cancer regulation[21,23]. This study used mRNA profiling and AI-based interpretation to investigate boric acid's effects on MYC-driven transcription and caspase-dependent apoptosis in breast cancer cells.

Materials and Methods

Cell culture

MCF-7 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) with 1 g/L D-Glucose, 4 mM L-Glutamine, and 1 mM Pyruvate (Gibco, USA), supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. MDA-MB-231 cells were cultured in DMEM/F-12 with 2.5 mM L-Glutamine and 15 mM HEPES (Gibco, USA), supplemented with 10% FBS and 1% penicillin-streptomycin

Assessment for the survival of cells

The cell viability test was used to assess the impacts of various concentrations of B at 0.1, 1, 10, 50, 100, and 1000 ng/mL on cells after 24 and 72 hours. Here 5×10^3 cells of the twentieth passage were treated. MCF-7 and MDA-MB-231 cells were treated with B at the indicated concentrations for 24 h and 72 h independently. At each time point, cells were treated continuously from the time of seeding until the designated endpoint (24 h or 72 h). Untreated cells served as the control group. At the end of each incubation period, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) reagent was added to each well and incubated for an additional 2 hours.

mRNA Expressions

Total RNA was isolated from MCF-7 and MDA-MB-231 cells at 24 and 72 hours using a monophasic solution of phenol and guanidine isothiocyanate (TRIzol Reagent; Invitrogen, Camarillo, CA, USA) according to the manufacturer's instructions. Complementary DNA (cDNA) synthesis was performed using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA) according to the manufacturer's protocol. The mRNA expression levels of ENDOG, Caspase-3, Caspase-8, Caspase-9, BAX, BCL-2, FAS, FASLG, MYC, and GAPDH were quantified by quantitative real-time PCR (qPCR) using Maxima SYBR Green qPCR Master Mix (2X)

(Thermo Scientific, Waltham, MA, USA). Gene-specific primer sequences are listed in Table 1. Reactions were performed in a total volume of 20 µL. The thermal cycling conditions were as follows: initial denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s, annealing at 60°C for 30 s, and extension at 72°C for 30 s. All reactions were performed using a BIORAD-CFX real-time PCR system (USA). Relative mRNA expression levels were calculated using the 2–ΔΔCt method, with GAPDH serving as the reference gene for normalization. All samples were run in triplicate.

Table 1. Human-origin and apoptosis-related gene primer sequences, 5'-3'

Primers	**Forward (5'-3')**	**Reverse (3'-5')**
ENDOG	GCATGCCTGGAACAACCTTGAGAA	TGCCTCCAGGATCAACACCTTGAA
Caspase3	TTAATAAAGGTATCCTCGGGGT	CATGGAGAACACTCCCAGCCT
Caspase8	TCTGGAGCATCTGCTGTCTG	CCTGCCTGGTGTCTGAAGTT
Caspase9	GGCTGTCTAC GGCAC AGATG GA	CTGGCTCGGGGTTACTGCCAG
BAX	AGTGGCAGCTGACATGTTTT	GGAGGAAGTCCAATGTCCAG
BCL-2	CCGGGAGATCGTGATGAAGT	ATCCCAGCCTCCGTTATCCT
FAS	GCCAATTCTGCCA! TAAGC	TTGGCTTCATTGACACCACTCGGGGT
FASLG	GGATTGGGCCTGGGGATGTTTCA	TTGTGGCTCAGGGGCAGGTTGTTG
MYC	TGA GGA GAC ACC GCCCAC	CAACATCGATTTCTTCCTCATCTT C3
GAPDH	CAAGGTCATCCATGACAACTTTG	GTCCACCACCCTGTTGCTGTAG

Ethical Statement

The experiments used established human breast cancer cell lines (MCF-7 and MDA-MB-231). With no primary tissues, patient data, or animal models used, the study does not require ethics committee approval.

Statistical Analysis

Tukey's honestly significant difference (HSD) test and Dunnett's test were employed in a one-way analysis of variance (ANOVA) to investigate the RT-PCR and cell viability results.

Extreme Gradient Boosting (XGBoost) Model Development

To systematically evaluate the non-linear impacts of boric acid (B) on cellular dynamics, two independent machine learning frameworks were established using the Extreme Gradient Boosting (XGBoost) regressor algorithm. XGBoost is a scalable tree-boosting system that implements gradient-descent optimization architectures optimized for multi-dimensional biological datasets. XGBoost Model 1 (Cell Viability Predictor) was developed to map experimental exposure parameters directly to cell viability percentages. XGBoost Model 2 (mRNA Expression Predictor) was designed to predict the relative transcription fold-changes of core apoptotic regulatory pathways. The models' parameters are shown in Table 2.

Table 2. Input and output limitations of developed XGBoost Models

	Inputs				**Output**
XGBoost Model 1	Time	Cell Types	Concentrations		Cell Viability
XGBoost Model 2	Time	Cell Types	Gene	Concentrations	mRNA Expression

The experimental input domains and targeting profiles for both models were explicitly formalized to map architectural parameters into quantitative target outcomes. Feature Set for Model 1 includes Time (24 h and 72 h; continuous numerical representation), B (0.1, 1, 10, 50, 100, and 1000 ng/mL; numerical scaling), and Cell Type (MCF-7 and MDA-MB-231). Feature Set for Model 2 includes Time (24 h and 72 h), B (1, 10, 50, and 100 ng/mL), Cell Type, and Gene Name (categorical spectrum encompassing ENDOG, caspase3, caspase8, caspase9, BAX, BCL-2, FAS, FASLG, and MYC). To facilitate optimal multi-variant algorithmic parsing, categorical features (Cell Type and Gene Name) were mapped into numerical vector spaces using one-hot encoding frameworks prior to model deployment. This avoids imposing artificial ordinal constraints on biological classes.

To prevent programmatic overfitting while maximizing generalized testing performance on non-homogeneous profiles, the collective experimental data matrix was partitioned into an 80% training repository and a 20% independent validation testing partition. Hyperparameter optimization was executed through an exhaustive grid-search pipeline integrated with internal k-fold cross-validation (k=5) on the training cohort. The optimized hyperparameter optimization matrix encompassed the following operational variations:

- Learning Rate: Explored within the bounds of [0.01, 0.05, 0.1, 0.2] to systematically scale step-size shrinkage.

- Maximum Tree Depth (max_depth): Varied from 3 to 8 to effectively regulate tree architectural complexity and mitigate model variance.

- Number of Estimators (n_estimators): Investigated from 50 to 500 steps to track structural boosting stabilization points.

Subsample weights and column sampling metrics were closely controlled to ensure tree construction steps were regularized against data-density imbalances. To quantify the contribution of each variable, feature importance was calculated, identifying Boric Acid Concentration (0.53163) as the primary driver of model predictions, followed by Cell Type (0.235684) and Time (0.232685). The performance of the models derived by both approaches has largely been evaluated with the Margin of Deviation (MoD) which determines the percentage difference between the predicted values (result from the creation of the XGBoost models) based on actual target values obtained in the experiment. It expresses Equation 1, the key to the MoD calculation:

$$MoD\ (\%) = \left[\frac{X_{targ} - X_{pred}}{X_{targ}}\right] x\ 100 \qquad (1)$$

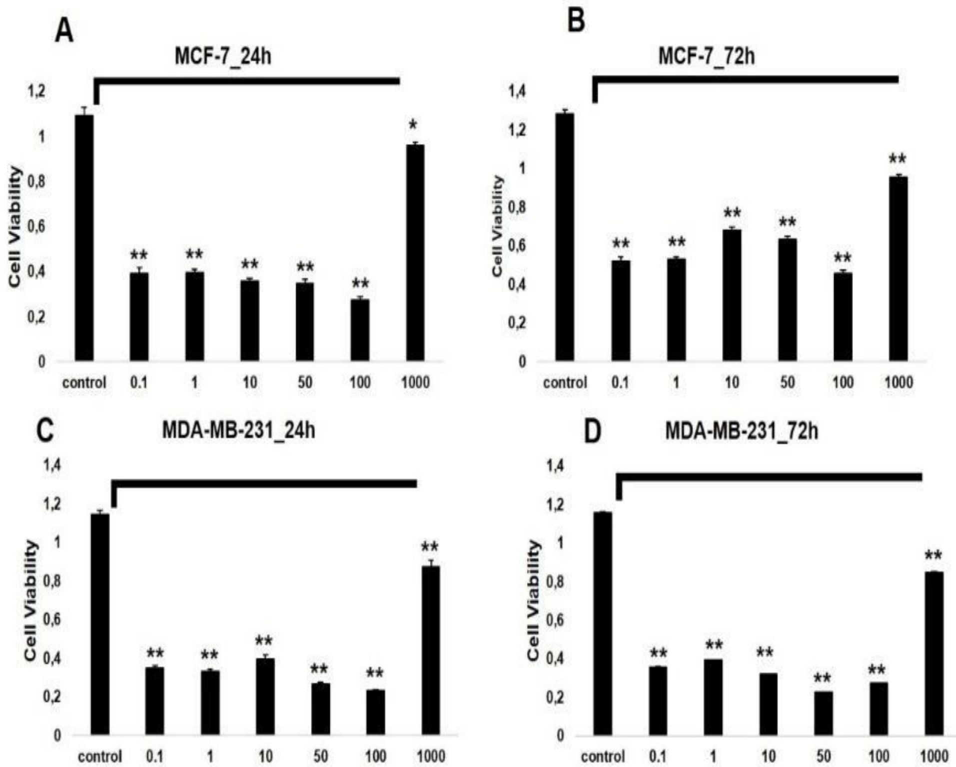
Here, X_{targ} represents the actual target value obtained from the experimental measurements (cell viability or mRNA expression), and X_{pred} represents the corresponding value predicted by the developed XGBoost model. The MoD is helpful in determining the predictive reliability and error characteristics of the AI models compared to biological changes.

Results

Boric acid diminished the survival rate of breast cancer cells.

Boric acid reduced MCF-7 and MDA-MB-231 breast cancer cell viability at 24 h and 72 h. MCF-7 viability decreased dose-dependently up to 1000 ng/mL at 24 h ($p<0.01$) and at 72 h with 0.1-100 ng/mL, peaking at 100 ng/mL ($p<0.01$). MDA-MB-231 viability decreased with 0.1-100 ng/mL at both times ($p<0.01$), strongest at 50-100 ng/mL. MDA-MB-231 cells were more sensitive, indicating greater triple-negative phenotype susceptibility.

Figure 1. Boric acid effects on MCF-7 and MDA-MB-231 breast cancer viability. MCF-7 cells were treated with boric acid (0.1-1000 ng/mL) for 24 h and 72 h (**A**-**B**); ***p < 0.001 vs control. MDA-MB-231 cells received equivalent treatments at 24 h and 72 h (**C**-**D**) ; ***p < 0.001 vs control.



mRNA Expression Results

Boric acid's impacts on apoptosis were investigated via intrinsic and extrinsic markers in breast cancer cells. Both lines showed increased ENDOG expression after treatment, with MCF-7 increases at 72h (100 ng/mL) and MDA-MB-231 increases at 72h (50-100 ng/mL) (p<0.01) (Figure 2A-D). Caspase genes increased in both lines (Figure 3-5). MCF-7 showed increased caspase-3 at 72 h with 100 ng/mL and at 50-100 ng/mL at 24 h (Figure 3A, 3B), while M(p < 0.01)DA-MB-231 peaked at 50 ng/mL after 24 h (Figure 3C, 3D). Caspase-8 and -9 increased dose-dependently (p<0.01) (Figure 4, 5). Pro-apoptotic BAX increased dose-dependently (p<0.01) (Figure 6A-D), while Bcl-2 decreased in both lines (Figure 7A-D). FAS/FASLG increased significantly (p<0.01) (Figure 8A-D, 9A-D). MYC expression increased in both lines (p<0.01) (Figure 10A-D). Boric acid activated both extrinsic and intrinsic apoptotic pathways.

Figure 2. Effect of B on ENDOG gene mRNA expression in MCF-7 at 24h (**A**), 72h (**B**), and MDA-MB-231 at 24h (**C**), 72h (**D**). B concentrations (1, 10, 50, 100 ng/mL) were tested using quantitative reverse transcription-polymerase chain reaction (RT-PCR). Bars show mean with 95% confidence interval. *p<0.05, **p<0.01 indicate significant differences versus control.

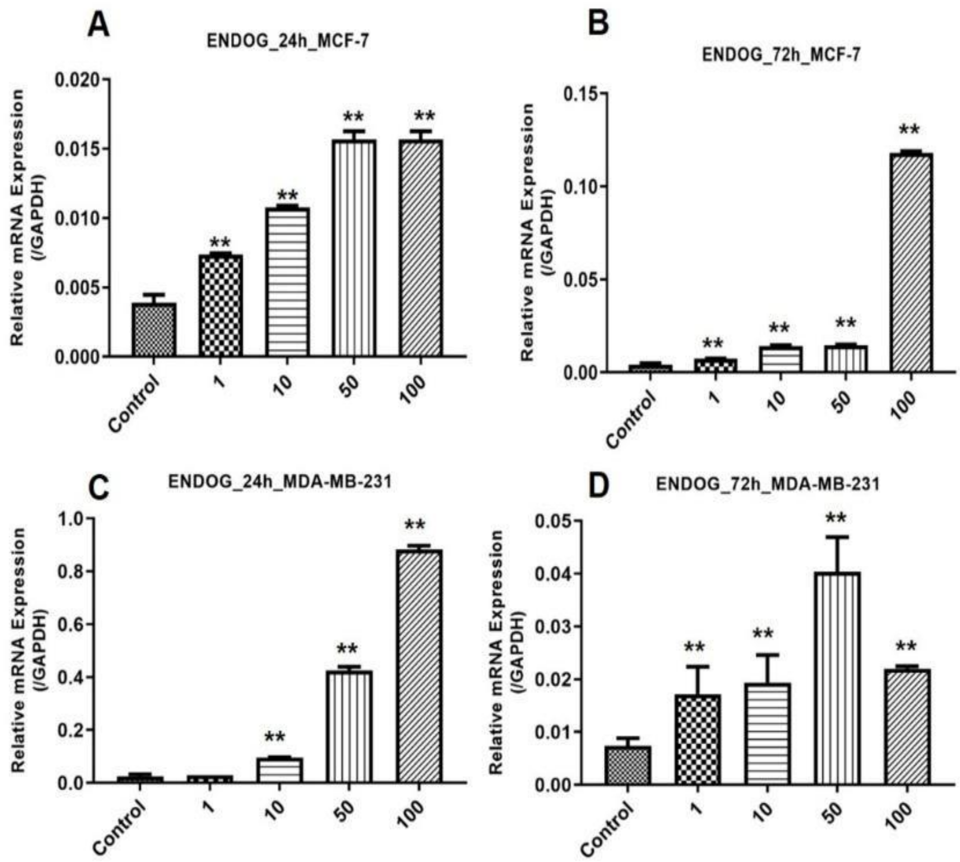


Figure 3. The effect of B on caspase-3 expression in MCF-7 at 24h (**A**), 72h (**B**) and MDA-MB-231 at 24h (**C**), 72h (**D**). B concentrations (1, 10, 50, 100 ng/mL) were tested on caspase-3 mRNA expression in MCF-7 and MDA-MB-231 using RT-PCR at 24h and 72h. Bars show mean with 95% CI. *p<0.05, **p<0.01 vs control.

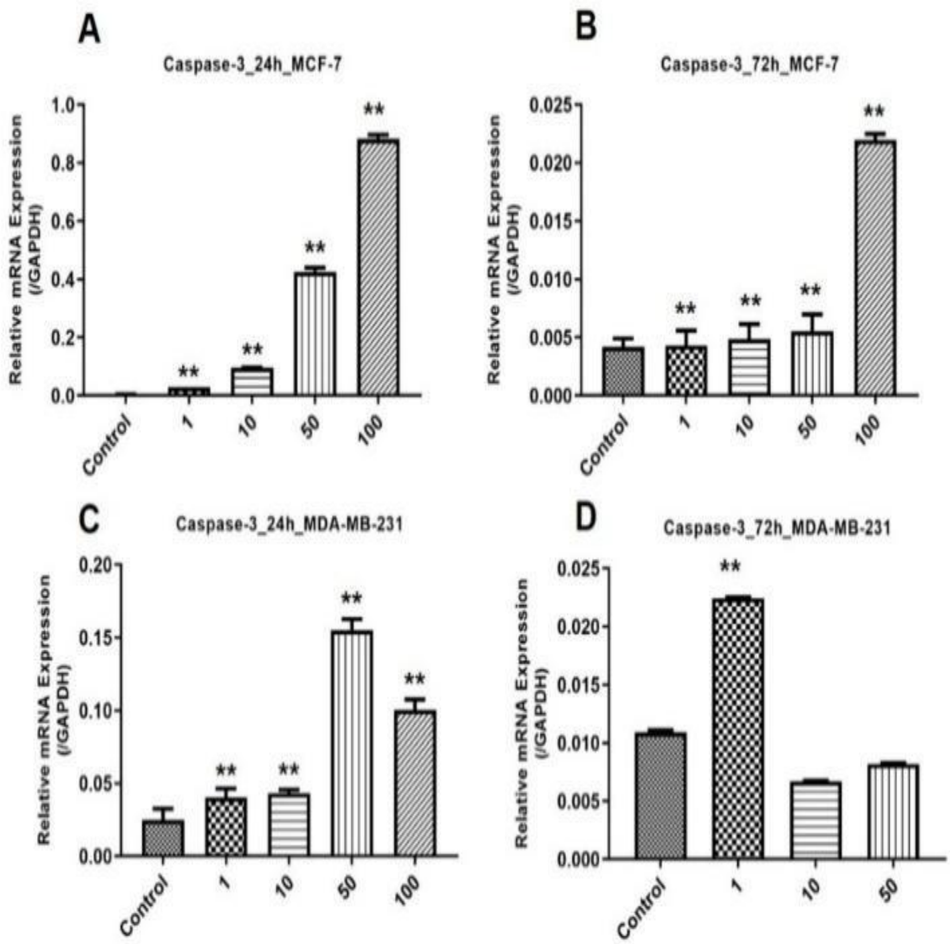
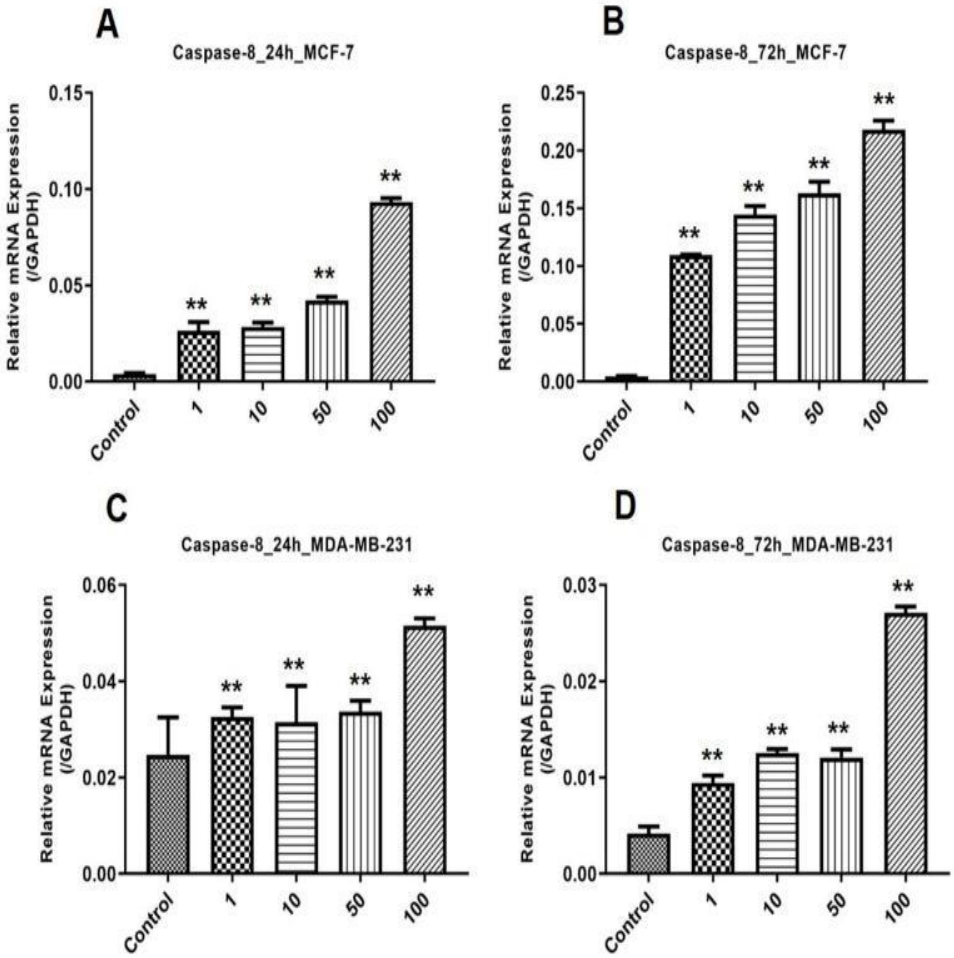


Figure 4. Effect of B on caspase-8 expression in MCF-7 at 24 h (**A**), 72h (**B**) and MDA-MB-231 at 24 h(**C**), 72h (**D**). B concentrations (1, 10, 50, 100 ng/mL) were tested on caspase-8 mRNA expression in MCF-7 and MDA-MB-231 using RT-PCR at 24 h and 72 h. Bars show mean with 95% confidence interval. *p<0.05, **p<0.01 vs control.



Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Figure 5. B affecting caspase-9 expression in MCF-7 at 24 h (**A**), 72h (**B**), and MDA-MB-231 at 24 h (**C**), 72 h (**D**). Effect of B concentrations (1, 10, 50, 100 ng/mL) on Caspase-9 gene mRNA expression in MCF-7 and MDA-MB-231 using RT-PCR at 24 h and 72 h, with bars showing mean at 95% confidence interval. *p<0.05, **p < 0.01 versus control.

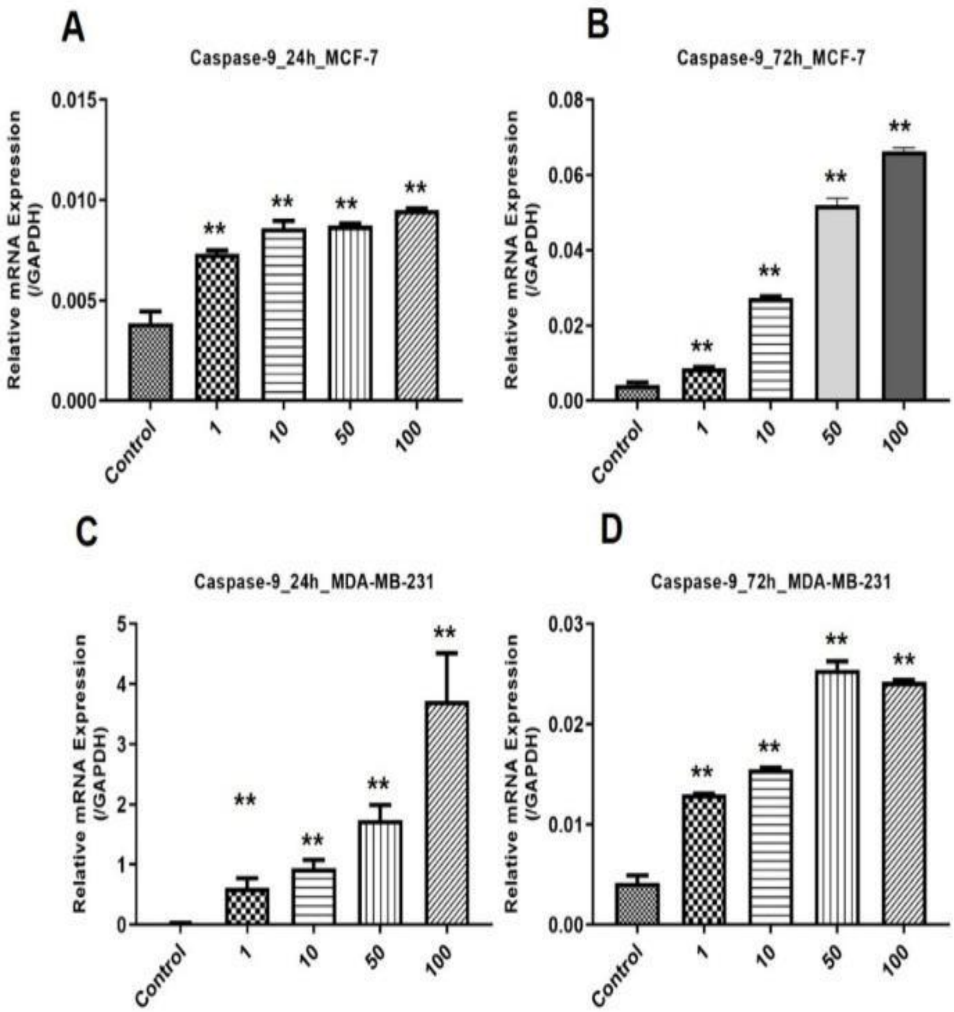


Figure 6. Effects of B on BAX expression in MCF-7 at 24h (**A**), 72h (**B**) and MDA-MB-231 at 24 h(**C**), 72h (**D**). B concentrations (1, 10, 50, 100 ng/mL) effects on BAX gene mRNA expression in MCF-7 and MDA-MB-231 using RT-PCR at 24 h and 72 h. Bars show mean with 95% confidence interval. *p<0.05, **p<0.01 vs control.

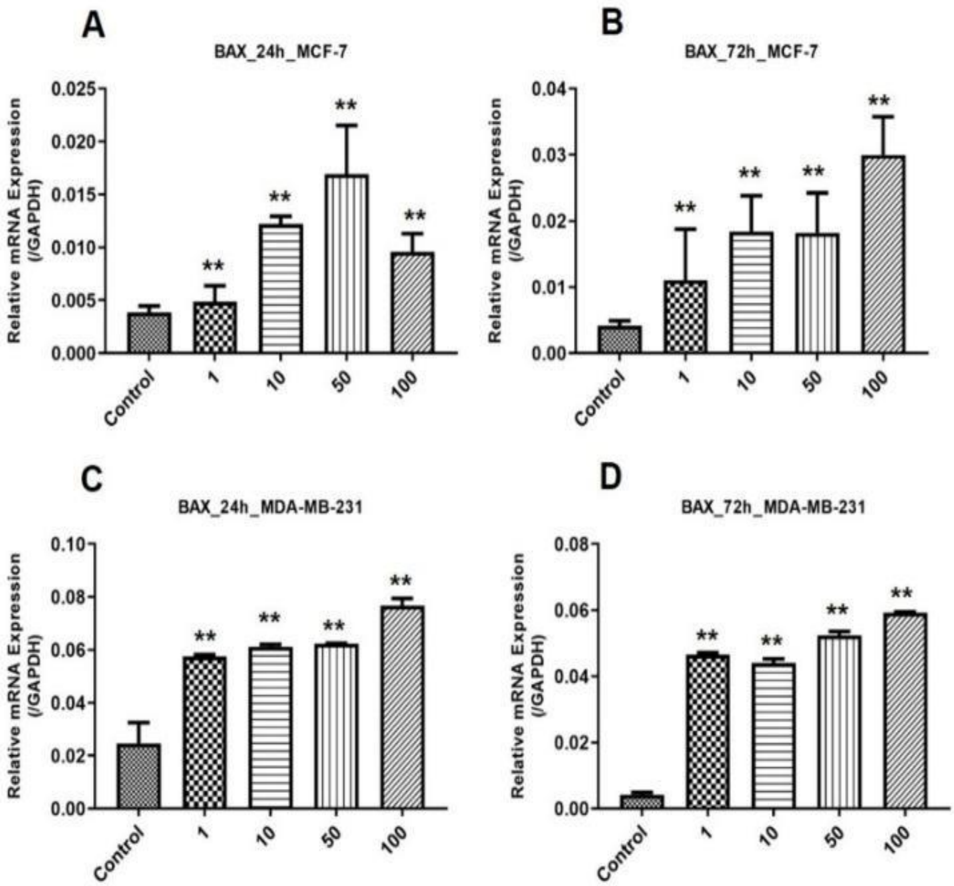


Figure 7. Influence of B on Bcl-2 expression in MCF-7 at 24 h (**A**), 72h (**B**) and MDA-MB-231 at 24 h (**C**), 72 h (**D**). Effect of B concentrations (1, 10, 50, 100 ng/mL) on Bcl-2 mRNA expression in MCF-7 and MDA-MB-231 using RT-PCR at 24 h and 72 h; bars show mean with 95% confidence interval. *p<0.05, **p<0.01 vs control.

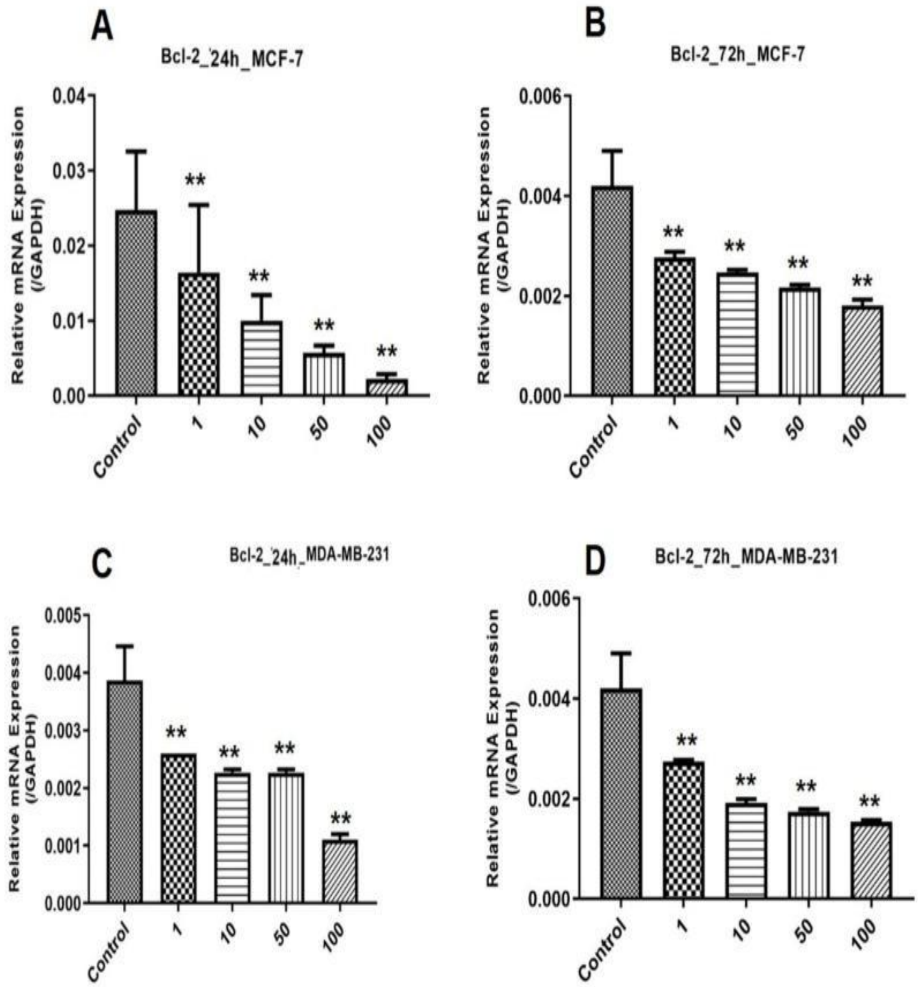


Figure 8. Effect of B on FAS expression in MCF-7 at 24 h (**A**), 72h (**B**) and MDA-MB-231 at 24 h (**C**), 72h (**D**). B concentrations (1, 10, 50, 100 ng/mL) were tested on FAS gene mRNA expression in MCF-7 and MDA-MB-231 using RT-PCR at 24 h and 72 h. Bars show mean with 95% confidence interval. *p<0.05, **p<0.01 vs control.

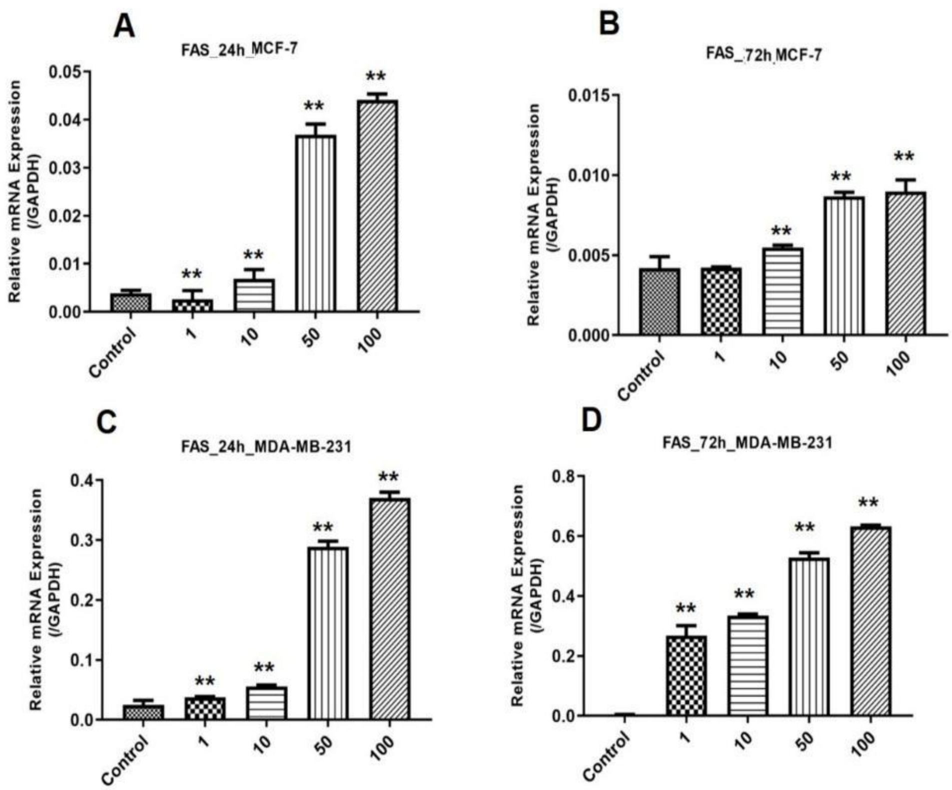
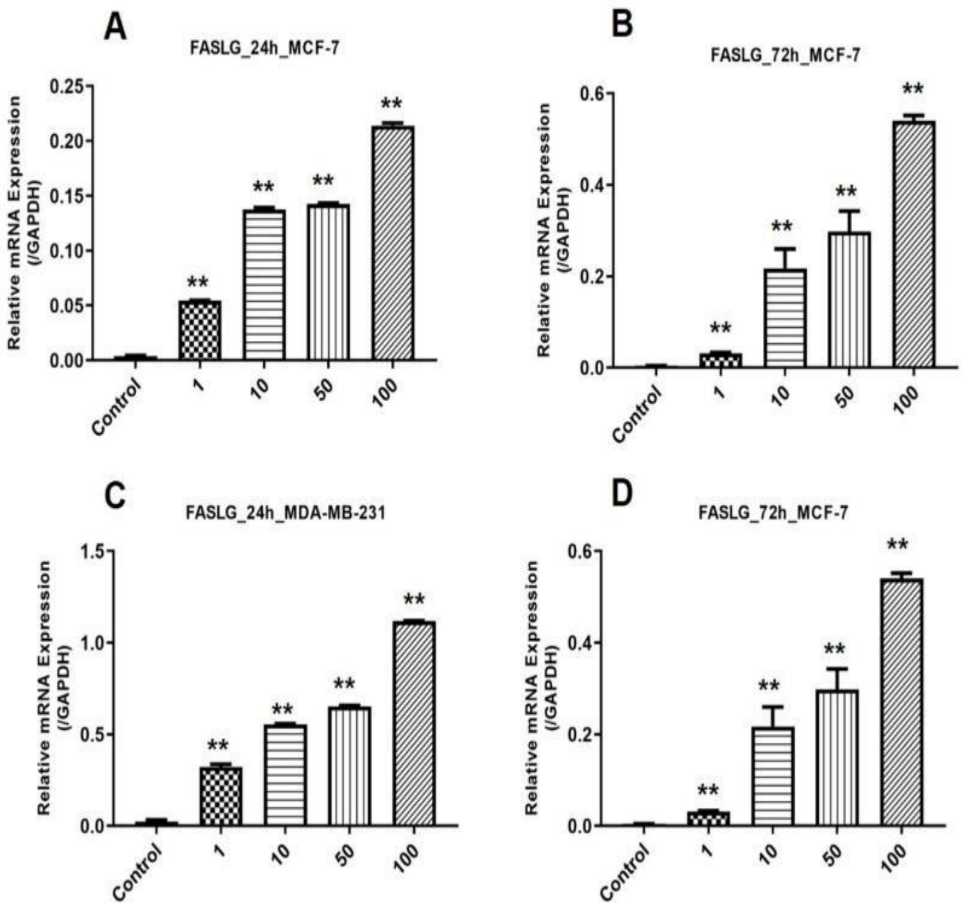
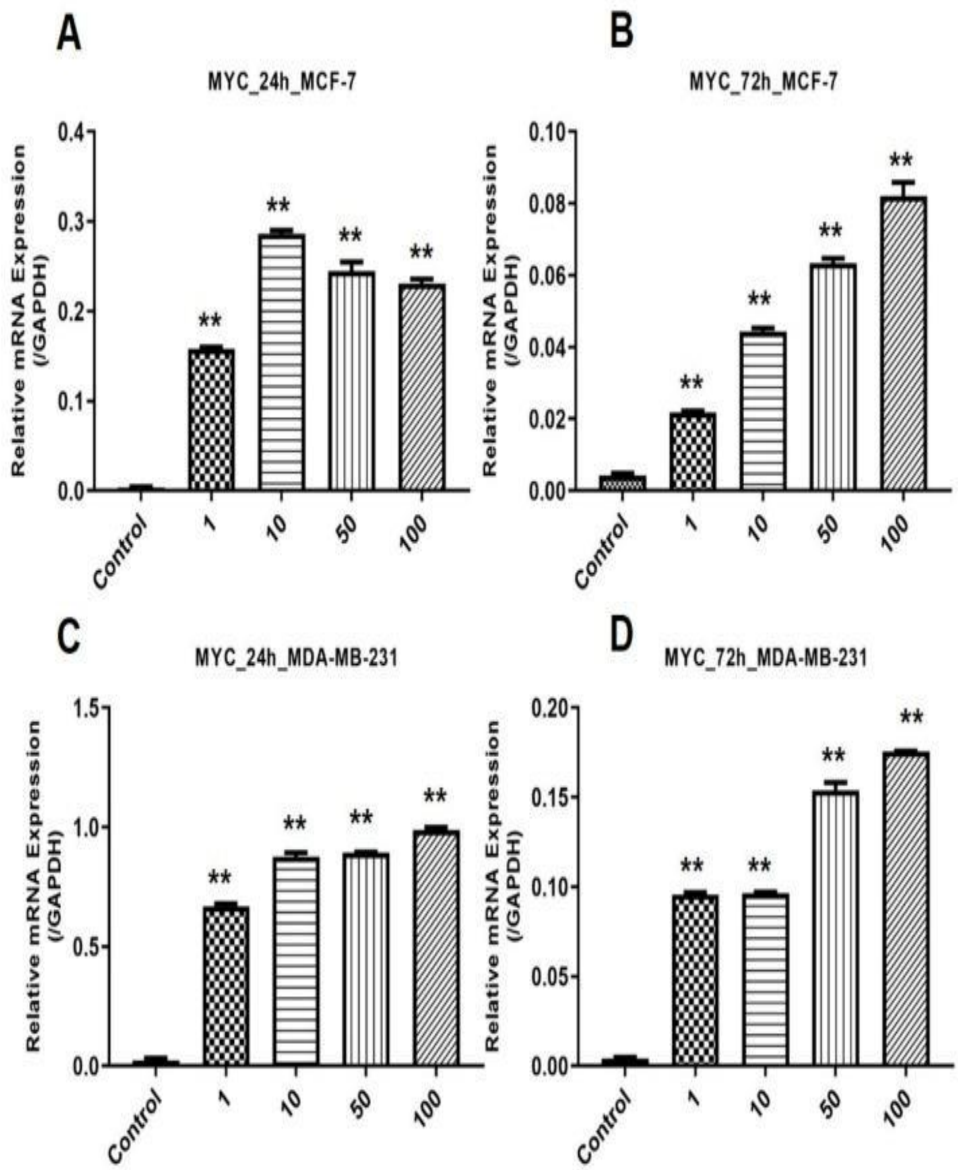


Figure 9. Effect of B on FASLG expression in MCF-7 at 24 h (**A**), 72h (**B**) and MDA-MB-231 at 24 h(**C**), 72h (**D**). B concentrations (1, 10, 50, 100 ng/mL) were tested on FASLG gene expression in MCF-7 and MDA-MB-231 using RT-PCR at 24 h and 72 h. Bars show mean with 95% confidence interval. *p<0.05, **p<0.01 vs control.



Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Figure 10. B effect on MYC expression in MCF-7 at 24 h (**A**), 72 h (**B**) and MDA-MB-231 at 24 h (**C**), 72 h (**D**). B concentrations (1, 10, 50, 100 ng/mL) affected MYC gene mRNA expression in MCF-7 and MDA-MB-231 cells at 24 h and 72 h by RT-PCR. Bars show mean with 95% confidence interval. *p<0.05, **p<0.01 vs control.



XGBoost Model Results

There are significant limitations in mapping boric acid to two optimized XGBoost models specifically designed to predict breast cancer cell viability (Model 1) and apoptosis mRNA expression (Model 2).

XGBoost Model 1: Cell Viability Prediction

Two optimized XGBoost models were used to predict boric acid's effect on breast cancer cell viability (Model 1) and apoptosis mRNA expression (Model 2). Data showed the model based on in vitro results was inadequate for revealing boric acid's effect. Figure 11A compares experimental and predicted cell viability for XGBoost Model 1. While values show a linear trend, indicating a general match, prediction accuracy varies significantly. Boric acid-induced cytotoxicity is undetectable from the total mass using Model 1 due to high MoD in Equation 1, with prediction errors reaching ±100%. Similarly, comparative analysis of Cell Viability (Figure 13A) shows the absolute error between actual and predicted values.e predicted) suggests the model consistently diverges by ±0.2 units (bad for prediction performance).

XGBoost Model 2: mRNA Expression Prediction

XGBoost Model 2's performance in predicting mRNA expression of apoptosis genes showed poor stability. PCR results were compared to XGBoost predictions (Figure 11B), with Figure 12B showing high variability in prediction accuracy. MoD values for mRNA Expression (Equation 1) ranged from -4000% to +2000%, indicating substantial deviations from experimental values. Due to this large error distribution, the model failed to accurately predict molecular stress response and gene expression changes from boric acid treatment. Both XGBoost models, despite optimization, could not predict cell viability and mRNA expression in MCF-7 and MDA-MB-231 cell lines.

Figure 11. Experimental target values were compared with XGBoost model predictions: cell viability with Model 1 predictions (**A**), and target mRNA expression with Model 2 predictions (**B**).

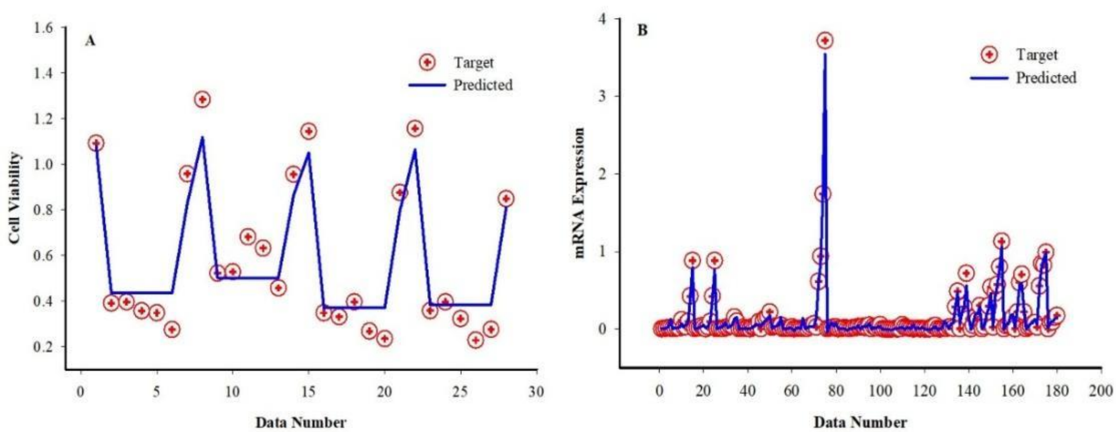


Figure 12. Ministry of Defence examination of the anticipated values for the XGBoost models. MoD distribution for Model 1 (cell viability prediction) (**A**). Distribution of MoD for Model 2 (mRNA Expression Prediction) (**B**).

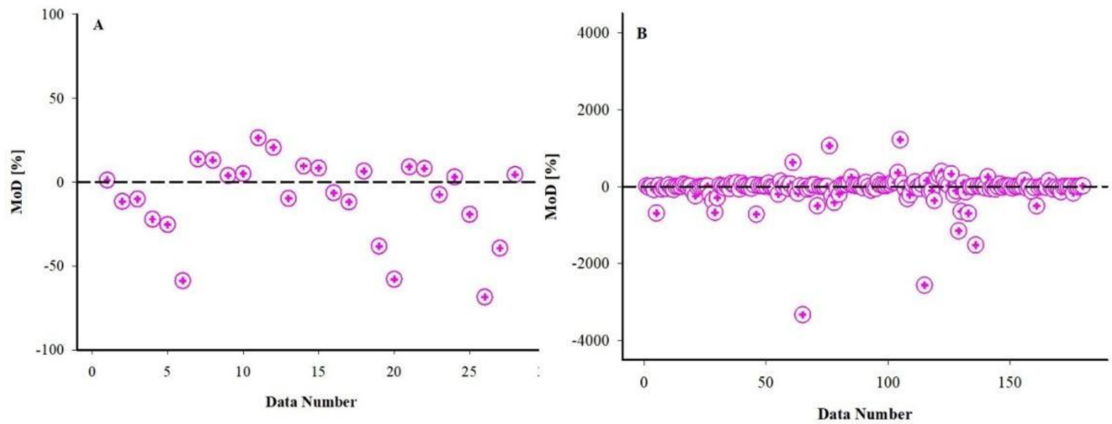
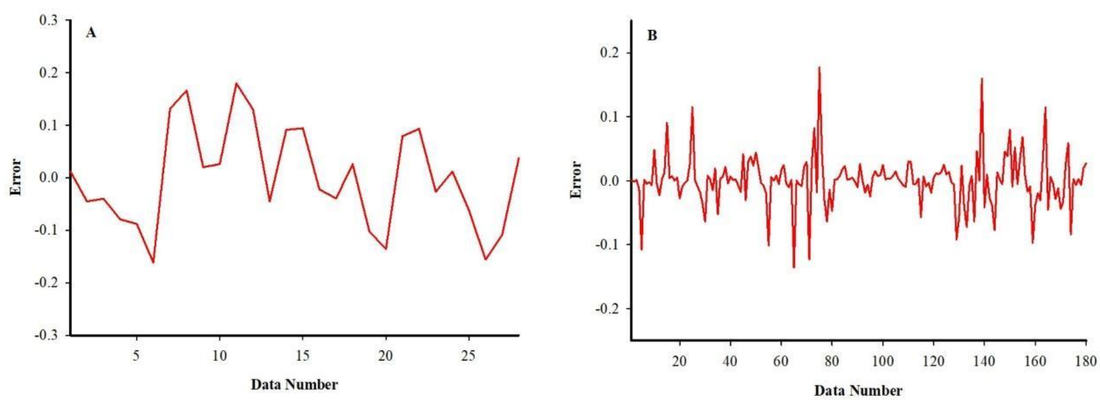


Figure 13. Comparison of expected values from XGBoost with experimental target values. Error distribution of cell viability prediction using XGBoost: Model 1 (**A**). Error distribution of mRNA expression prediction using XGBoost: Model2 (**B**).



Discussion

In the present study, B reduced cell viability in both MCF-7 and MDA-MB-231 cells in a concentration- and time-dependent manner. Notably, MDA-MB-231 cells exhibited markedly greater sensitivity to B at both 24 and 72 hours compared to MCF-7 cells, suggesting that the triple-negative phenotype—characterized by the absence of estrogen receptor, progesterone receptor, and HER2—confers heightened susceptibility to boric acid-induced cytotoxicity. This differential response is consistent with the greater aggressiveness and distinct metabolic profile of TNBC cells[24,25].

Research shows B has varying cytotoxic and apoptotic effects by cell type. HeLa cells show higher sensitivity than normal HUF cells, with lower IC50 value and G1 arrest, nuclear abnormalities, and autophagy-related gene changes linked to endoplasmic reticulum (ER) stress[24]. Studies of human breast tumor cell lines showed gene expression differences drive pro-apoptotic drug development[25]. Boric acid modulates both apoptosis pathway genes and genes essential for apoptosis. The upregulation of GRP78, p-IRE1α, p-PERK, CHOP stress pathways, and cleaved caspase-3 indicates ER stress links to programmed cell death. Cellular death pathways including p53, MAPK, TNF, PI3K/AKT, and NF-κB are modulated through defined factors[26]. Quercetin modifies apoptosis-associated gene expression by upregulating pro-apoptotic genes and downregulating anti-apoptotic genes, impacting death receptor signaling[27]. Artificial intelligence can analyze gene expression to understand interactions from boric acid exposure, identifying key regulators and forecasting cell responses to boric acid-activated stress in cancer cells. AI targets factors like caspase activation and unfolded protein response (UPR) that trigger apoptotic response[24]. Different cellular lineages demonstrate biological heterogeneity, with analyses showing boric acid targets various apoptotic pathways. The protective effect against apoptosis in mouse granulosa cells during heat stress indicates boron compounds modulate ER stress signals selectively[28]. While boric acid induces apoptosis in DU-145 prostate cancer cells[29], effects vary across cancers. It slows human breast cancer cells without consistently inducing apoptosis, while calcium fructoborate causes apoptosis. Different boron compounds function through varied approaches, with effectiveness depending on cell type. This research advances knowledge through studying endoplasmic reticulum-stress response and autophagy-related gene expression. Boric acid shows promise in breast cancer management, with studies showing both boric acid and calcium fructoborate inhibit proliferation in MDA-MB-231 cells dose-dependently.

Calcium fructoborate significantly induced apoptosis by reducing p53 and Bcl-2 protein levels and activating caspase-3[21]. A central finding of this study is the coordinated upregulation of BAX, ENDOG, Caspase-9, and Caspase-3 mRNA alongside downregulation of BCL-2, resulting in a pro-apoptotic shift in the BAX/BCL-2 ratio in both cell lines. This pattern is mechanistically consistent with mitochondria-mediated intrinsic apoptosis[26]. Notably, the upregulation of ENDOG—an endonuclease released from mitochondria during apoptosis—in both cell lines represents a finding not previously documented for boric acid in breast cancer cells, supporting caspase-independent nuclear DNA fragmentation as a concurrent mechanism.

The present data also demonstrate activation of the extrinsic apoptotic pathway, evidenced by elevated FAS, FASLG, and CASP8 mRNA levels. Importantly, this activation displayed a cell-type-specific temporal pattern: extrinsic pathway induction occurred at 72 hours in MCF-7 cells but as early as 24 hours in MDA-MB-231 cells. This earlier engagement of the death receptor pathway in MDA-MB-231 cells may partly explain their greater overall apoptotic susceptibility and suggests that the two pathways are not activated simultaneously but rather follow a cell-phenotype-dependent sequence.

Perhaps the most mechanistically distinctive finding of this study is the upregulation of MYC expression under B treatment in both cell lines, albeit with an opposing temporal pattern: MYC was induced at 24 hours in MDA-MB-231 cells and at 72 hours in MCF-7 cells. This is paradoxical given MYC's canonical role as an oncogene promoting proliferation; however, sustained or aberrant MYC overexpression in the context of concurrent pro-apoptotic signaling is well-recognized as a trigger for apoptosis—a phenomenon termed 'MYC-induced apoptosis' first demonstrated in fibroblasts[30,31] and subsequently linked to the CD95/FAS pathway when BCL-2 survival signals are suppressed [32]— particularly when survival signals such as BCL-2 are simultaneously suppressed, as observed here. This MYC-driven, caspase-dependent apoptotic axis may represent a novel mechanism of boric acid action in breast cancer.

While earlier studies suggested boric acid might inhibit breast cancer through pathways other than apoptosis[21], the present findings demonstrate unequivocally that boric acid induces both intrinsic and extrinsic apoptotic programs in MCF-7 and MDA-MB-231 cells. Beyond the molecular mechanisms identified here, the translational potential of B in breast cancer management has been further explored through nanomedicine-based delivery strategies.

Nanomedicine shows promise for delivering boric acid to breast cancer. Boric acid and paclitaxel in polymeric nanocarriers showed higher drug uptake and potentially more potent effects. Nanoparticle systems with boric acid present a novel strategy for treating breast cancer, addressing drug resistance and side effects[33]. Understanding molecular aspects of breast cancer stages is crucial for personalizing therapy. Metabolomics and nanoparticle-assisted therapy improve cancer detection[34,35]. LAT1 shows upregulated expression associated with boron-containing compounds[36], which accumulate more in aggressive breast cancers[37]. This enables targeted boron injection for neutron capture therapy[38]. Boric acid-based therapies within personalized regimens offer promising solutions[21], particularly when combined with other treatments based on tumor biology and patient genetics[22]. The failure of the XGBoost models to predict mRNA expression outcomes is itself a biologically informative finding: it quantitatively demonstrates that the regulatory network activated by boric acid—encompassing MYC-driven transcription, dual-pathway apoptosis, and cell-type-specific timing—cannot be captured by a feature set limited to time, concentration, and cell identity alone. This reinforces the conclusion that boric acid's mechanism is inherently multi-dimensional. The AI study aimed to develop models interpreting complex datasets to predict biological outcomes. While XGBoost Model 1 showed good universal fit, it produced large prediction errors upon detailed analysis. Specifically, in detail in Figure 12A the MoD for cell viability can often be seen as high percentages, showing that the model was unable to accurately predict the exact magnitude of cytotoxicity across different cell lines and at different concentrations. The failure was particularly pronounced for the XGBoost Model 2 (mRNA Expression) as the MoD (Figure 12B) shows pronounced variances, ranging from approximately -4000% to +2000%. The large deviation rates in other variables are also demonstrated with strong error distributions in Figures 13A and B in the above plot, suggesting that the models fail to adequately capture the quantitative properties of the biological response. That the optimized XGBoost models do not provide a good prediction of these results indicates that the settings available (e.g., temporal range, cell type, concentration, and gene names) were insufficient for capturing the complexity behind the biological responses due only to boric acid. The experimental results showed that boric acid induces elaborate biological processes: the stimulation of intrinsic and extrinsic apoptotic pathways occurs concurrently; it triggers phenotype-specific reaction in MDA-MB-231 cells, and delayed or late-phase activation in MCF-7 cells followed by high dosage and time dependence. Relative mRNA expression, as measured by the $2^{-\Delta\Delta Ct}$ approach, is the product of deep cross-talk within a vast number of signaling networks, including MYC-regulated transcription and caspase cascades, a dynamic cellular condition. Thus, the fall from grace of the evolved XGBoost models demonstrate one of the difficulties introduced by the reliance on typical experimental parameters (time, concentration, cell type) to model the high-dimensional dynamic regulatory systems of biological systems. Although AI has a considerable advantage in combining multiple types of data used for diagnosis and targeted therapies, accurate prediction of complex biological results, such as mRNA expression, requires a denser and wider feature set in addition to potentially multiomics data types (e.g., proteomics or epigenetics) that move beyond the scope of the main variables used in training the model.

Conclusion

Results show boric acid modulates mRNA expression for cell viability and apoptosis in breast cancer. AI enables pathway analysis and improves diagnosis through imaging and biomarkers. AI models predict outcomes and accelerate drug development. This study advances therapeutics while showing AI's role in cancer biology. However, XGBoost models showed ±100% errors in Cell Viability and -4000% to +2000% instability in mRNA expression, requiring more data.

REFERENCES

1. Tang L, Mo D, Wei D, et al. Long non-coding RNA MIR200CHG promotes breast cancer proliferation, invasion, and drug resistance by interacting with and stabilizing YB-1. *NPJ Breast Cancer*. 2021;7(1):1–12.

2. Antoon JW, White MD, Slaughter EM, et al. Targeting NFκB mediated breast cancer chemoresistance through selective inhibition of sphingosine kinase-2. *Cancer Biol Ther*. 2011;11(7):678–689.

3. Szostakowska M, Trębińska-Stryjewska A, Grzybowska EA, Fabisiewicz A. Resistance to endocrine therapy in breast cancer: molecular mechanisms and future goals. *Breast Cancer Res Treat*. 2018;173(3):489–497.

4. Li Y, Li Z. Potential mechanism underlying the role of mitochondria in breast cancer drug resistance and its related treatment prospects. *Front Oncol*. 2021;11:629614.

5. Wilson T, Johnston P, Longley D. Anti-apoptotic mechanisms of drug resistance in cancer. *Curr Cancer Drug Targets*. 2009;9(3):307–319.

6. Kveiborg M, Rank F, Wewer UM, et al. A role for ADAM12 in breast tumor progression and stromal cell apoptosis. *Cancer Res*. 2005;65(11):4754–4761.

7. Maniam S, Maniam S. Small molecules targeting programmed cell death in breast cancer cells. *Int J Mol Sci*. 2021;22(18):9722.

8. Yuan L, Cai Y, Zhang L, Liu S, Li P, Li X. Promoting apoptosis, a promising way to treat breast cancer with natural products: a comprehensive review. *Front Pharmacol*. 2021;12:801662.

9. Bhinder B, Elemento O, Madhukar NS, Gilvary C. Artificial intelligence in cancer research and precision medicine. *Cancer Discov*. 2021;11(4):900–915.

10. Fountzilas E, Pearce T, Baysal MA, Chakraborty A, Tsimberidou AM. Convergence of evolving artificial intelligence and machine learning techniques in precision oncology. *NPJ Digit Med*. 2025;8(1):1–10.

11. Torrente M, Sousa PA, Hernández R, et al. An artificial intelligence-based tool for data analysis and prognosis in cancer patients: results from the Clarify study. *Cancers (Basel)*. 2022;14(16):4041.

12. Garg P, Singhal G, Kulkarni P, Horne D, Salgia R, Singhal SS. Artificial intelligence-driven computational approaches in the development of anticancer drugs. *Cancers (Basel)*. 2024;16(22):3884.

13. Tanoli Z, Vähä-Koskela M, Aittokallio T. Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opin Drug Discov*. 2021;16(9):977–989.

14. Momoli C, Costa B, Lenti L, et al. The evolution of anticancer 3D in vitro models: the potential role of machine learning and AI in the next generation of animal-free experiments. *Cancers (Basel)*. 2025;17(4):700.

15. Huhulea EN, Huang L, Eng S, et al. Artificial intelligence advancements in oncology: a review of current trends and future directions. *Biomedicines*. 2025;13(4):951.

16. Nafissi N. The application of artificial intelligence in breast cancer. *Eurasian J Med Oncol*. 2024;8(3):235–244.

17. Bozkurt SB, Nielsen FH, Hakki SS. Boric acid reverses nicotine-induced cytokine expressions of human gingival fibroblasts. *Biol Trace Elem Res*. 2023;201(3):1174–1180.

18. Hakki SS, Bozkurt SB, Hakki EE. Boron regulates mineralized tissue-associated proteins in osteoblasts (MC3T3-E1). *J Trace Elem Med Biol*. 2010;24(4):243–250.

19. Hakki SS, Bozkurt SB, Hakki EE, Nielsen FH. Boron as boric acid induces mRNA expression of the differentiation factor tuftelin in pre-osteoblastic MC3T3-E1 cells. *Biol Trace Elem Res*. 2021;199(3):1534–1543.

20. Bozkurt SB, Hakki SS, Nielsen FH. Boric acid alleviates periodontal inflammation induced by IL-1β in human gingival fibroblasts. *J Trace Elem Med Biol*. 2024;84:127466

21. Scorei R, Ciubar R, Cimpean A, Ciofrangeanu CM, Mitran V, Iordachescu D. Comparative effects of boric acid and calcium fructoborate on breast cancer cells. *Biol Trace Elem Res*. 2008;122(3):197–205.

22. Barranco WT, Eckhert CD. Cellular changes in boric acid-treated DU-145 prostate cancer cells. *Br J Cancer*. 2006;94(6):884–890.

23. Kaya S, Doğan SY, Solak EK, Celep AGS, Demirkaya A. Designing of polymeric nanoparticles for enhanced breast cancer therapy: combining paclitaxel, boric acid and tannic acid for controlled drug delivery. *Chemistry Select*. 2024;9(7):e202304672.

24. Keyif B, Hacioglu C. Boric acid suppresses cell survival by triggering endoplasmic reticulum stress-induced autophagy in cervical cancers. *J Cell Mol Med*. 2025;29(14).

25. Messeha S, Mendonca P, Alwagdani H, Soliman K, Zarmouh N, Cotton C. Effects of gossypol on apoptosis-related gene expression in racially distinct triple-negative breast cancer cells. *Oncol Rep*. 2019;42(2).

26. Sun Y, Liu X, Jia X, Gao Q, Hou L. Involvement of apoptotic pathways in docosahexaenoic acid-induced benefit in prostate cancer: pathway-focused gene expression analysis using RT2 Profile PCR Array System. *Lipids Health Dis*. 2017;16(1).

27. Shang H, Lee C, Chen A, et al. Quercetin induced cell apoptosis and altered gene expression in AGS human gastric cancer cells. *Environ Toxicol*. 2018;33(11):1168–1181.

28. Xiong Y, Che C, He S, Yin Q, Jin E. Boron attenuates heat stress-induced apoptosis by inhibiting endoplasmic reticulum stress in mouse granulosa cells. *Biol Trace Elem Res*. 2020;199(2):611–621.

29. Hacioglu C, Kacar S, Kar F, Sahinturk V, Kanbak G. High concentrations of boric acid trigger concentration-dependent oxidative stress, apoptotic pathways and morphological alterations in DU-145 human prostate cancer cell line. *Biol Trace Elem Res*. 2019;193(2):400-409.

30. Li LY, Luo X, Wang X. Endonuclease G is an apoptotic DNase when released from mitochondria. *Nature*. 2001;412(6842):95-99.

31. Evan GI, Wyllie AH, Gilbert CS, Littlewood TD, Land H, Brooks M, Waters CM, Penn LZ, Hancock DC. Induction of apoptosis in fibroblasts by c-myc protein. *Cell*. 1992;69(1):119-28.

32. Hueber AO, Zörnig M, Lyon D, Suda T, Nagata S, Evan GI. Requirement for the CD95 receptor-ligand pathway in c-Myc-induced apoptosis. *Science*. 1997;278(5341):1305-9.

33. Gadag S, Nayak Y, Garg S, Nayak UY, Sinha S. Combination therapy and nanoparticulate systems: smart approaches for the effective treatment of breast cancer. *Pharmaceutics*. 2020;12(6):524.

34. Nehal N, Rohilla A, Sartaj A, Baboota S, Ali J. Folic acid modified precision nanocarriers: charting new frontiers in breast cancer management beyond conventional therapies. *J Drug Target*. 2024;32(8):855–873.

35. Perestrelo R, Luís C. Metabolomics in breast cancer: from biomarker discovery to personalized medicine. *Metabolites*. 2025;15(7):428.

36. Zheng X, Pan J, Lin D, Shao W. L-type amino acid transporter 1 in enhancing boron neutron capture therapy: mechanisms, challenges and future directions. *Int J Mol Med*. 2025;56(5):1–18.

37. Afzal MZ, Vahdat LT. Evolving management of breast cancer in the era of predictive biomarkers and precision medicine. *J Pers Med*. 2024;14(7):719.

38. Aristei C, Perrucci E, Alì E, et al. Personalization in modern radiation oncology: methods, results and pitfalls. *Front Oncol*. 2021;11.

Predicting Mortality and the Need for Early Intubation in Intensive Care Patients Using Machine Learning: A Pilot Study

Seval KILBASANLI[*], Andaç Batur ÇOLAK[**], Şerife Buket BOZKURT POLAT[***], Zeynep YÜKSEL TURHAN[****], Mustafa KAÇMAZ[*****], Seyyid Mehmet BULUT[******]

Abstract

Aim: This pilot study aimed to assess the feasibility and preliminary performance of machine learning models for predicting ICU mortality and the need for intubation within 48 hours using routinely collected admission data.

Method: Ten adult intensive care patients were included in this single-center, retrospective observational pilot study. For ANN development, 13 prespecified clinical and laboratory features per patient were restructured in long format, yielding 130 feature-level records. These records represented feature entries derived from 10 patients rather than independent patient-level observations. Demographic characteristics, primary diagnosis, APACHE II and SOFA scores, arterial blood gas parameters, oxygenation indicators (FiO_2, PaO_2/FiO_2), and routine metabolic–biochemical variables were analyzed. ML-based classification models predicted ICU mortality and the need for intubation within the first 48 hours.

Results: In the exploratory feature-level internal assessment, both the ICU mortality and early-intubation classifications yielded an AUC of 1.00, sensitivity of 100%, specificity of 100%, and an F1-score of 1.00. These estimates were derived from records originating from 10 patients and should be interpreted cautiously.

Conclusion: This pilot study demonstrates the feasibility of applying machine learning methods to routinely collected ICU admission data. The findings are preliminary and hypothesis-generating and should not be interpreted as established clinical predictive performance. Validation in larger, independent, prospective, and multicenter cohorts is required.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 02.02.2026 **Kabul / Accepted:** 20.07.2026
DOI: https://doi.org/10.38079/igusabder.1880038
[*] (Corresponding Author) Asst. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Anesthesiology and Reanimation, Niğde, Türkiye. E-mail: sevalkilbasanli@gmail.com ORCID https://orcid.org/0000-0003-4052-9593
[**] Assoc. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Computer and Information Sciences, Department of Information Systems and Technologies, Niğde, Türkiye. E-mail: bcolak@ohu.edu.tr ORCID https://orcid.org/0000-0001-9297-8134
[***] Assoc. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Medical Biochemistry, Niğde, Türkiye. E-mail: buketbozkurt@yahoo.com ORCID https://orcid.org/0000-0002-8641-2844
[****] Asst. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Anesthesiology and Reanimation, Niğde, Türkiye. E-mail: zeynepyuksel6@gmail.com ORCID https://orcid.org/0000-0002-8906-2149
[*****] Assoc. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Anesthesiology and Reanimation, Niğde, Türkiye. E-mail: muskac51@gmail.com ORCID https://orcid.org/0000-0002-8655-3882
[******] Res. Asst., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Medical Biochemistry, Niğde, Türkiye. E-mail: sbulut@ohu.edu.tr ORCID https://orcid.org/0000-0002-2010-7230

ETHICAL STATEMENT: The protocol was approved by the Non-Interventional Clinical Research Ethics Committee of Niğde Ömer Halisdemir University, Faculty of Medicine, Türkiye (Decision No: 2026/3, Date: 08.01.2026).

Keywords: Intensive care, machine learning, mortality, intubation, pilot study, decision support.

Yoğun Bakım Hastalarında Mortalite ve Erken Entübasyon Gereksiniminin Makine Öğrenmesi ile Öngörülmesi: Pilot Çalışma

Öz

Amaç: Bu pilot çalışmada, rutin yatış verilerine dayalı makine öğrenmesi modellerinin yoğun bakım mortalitesi ile ilk 48 saat içinde entübasyon gereksinimini öngörmedeki uygulanabilirliğinin ve ön performansının değerlendirilmesi amaçlandı.

Yöntem: Tek merkezli, retrospektif, gözlemsel pilot çalışmaya 10 erişkin yoğun bakım hastası dahil edildi. Yapay sinir ağı geliştirilmesi için her hastaya ait önceden belirlenmiş 13 klinik ve laboratuvar özelliği uzun veri formatında yeniden düzenlenerek toplam 130 özellik düzeyinde kayıt oluşturuldu. Bu kayıtlar, bağımsız hasta gözlemlerini değil, 10 hastadan türetilen özellik kayıtlarını temsil etmekteydi. Demografik özellikler, primer tanı, APACHE II ve SOFA skorları, arter kan gazı parametreleri, oksijenasyon göstergeleri (FiO_2, PaO_2/FiO_2) ve rutin metabolik–biyokimyasal değişkenler analiz edildi. Yoğun bakım mortalitesi ve ilk 48 saat içinde entübasyon gereksiniminin öngörülmesi amacıyla ML tabanlı sınıflandırma modelleri kullanıldı.
Bulgular: Özellik düzeyindeki keşifsel iç değerlendirmede, hem yoğun bakım mortalitesi hem de erken entübasyon sınıflandırması için AUC 1,00, duyarlılık %100, özgüllük %100 ve F1-skoru 1,00 olarak bulundu. Bu tahminler 10 hastadan türetilen kayıtlara dayandığından dikkatle yorumlanmalıdır.

Sonuç: Bu pilot çalışma, yoğun bakım yatışında rutin olarak elde edilen verilere makine öğrenmesi yöntemlerinin uygulanabilir olduğunu göstermektedir. Bulgular ön ve hipotez oluşturucu nitelikte olup, yerleşik bir klinik öngörü performansının kanıtı olarak değerlendirilmemelidir. Modellerin daha geniş, bağımsız, prospektif ve çok merkezli hasta gruplarında doğrulanması gereklidir.

Anahtar Sözcükler: Yoğun bakım, makine öğrenmesi, mortalite, entübasyon, pilot çalışma, karar destek sistemi.

Introduction

Intensive care units are clinical settings where life-threatening acute illnesses and multiple organ failures are managed and where the risk of mortality is high. Predicting mortality risk early on is valuable for planning the intensity of monitoring, timing interventions, and using resources effectively. In recent years, numerous studies and reviews have been published showing that machine learning (ML) approaches can be effective in predicting intensive care mortality[1–6].

The need for intubation is one of the most critical indicators of clinical deterioration in intensive care patients. Delayed intubation may be associated with difficulties in airway management, increased risk of complications, and poor clinical outcomes. Recently, externally validated studies and deep learning-based approaches have been reported to predict the need for intubation using interpretable ML models; models targeting time-dependent outcomes, such as the time to intubation during the first days of intensive care, have been developed[7–9].

Sepsis remains one of the most important predictors of intensive care mortality. Physiological deterioration in sepsis can progress rapidly, making early risk classification difficult. Approaches addressing imbalanced class problems have been shown to improve sepsis mortality prediction. Machine learning models have also been developed to predict mortality in patients with sepsis and in those with concomitant chronic kidney disease[10-12].

Although scores such as APACHE II and SOFA are widely used in clinical practice for mortality prediction, these scores have limitations in terms of early and accurate prediction at the individual level. In contrast, it has been reported that easily accessible variables such as electrolyte imbalances and routine laboratory parameters can emerge as meaningful predictors in ML-based models[13].

Intensive care patients on mechanical ventilation have a higher risk of mortality. ML models predicting hospital mortality and community approaches predicting prolonged ventilation requirements have been reported in this group. Studies analyzing factors associated with invasive mechanical ventilation using ML and models focusing on early mortality prediction in ARDS reveal significant potential for clinical decision support systems. A recent overview of artificial intelligence applications in mechanical ventilation emphasizes that decision support systems working with routine clinical data will become increasingly prominent[14-18].

In this context, alongside models developed with large datasets, it is important that small-sample pilot studies conducted with variables accessible in routine clinical practice demonstrate feasibility and lay the methodological groundwork for larger studies. Furthermore, it has been reported that not only mortality but also outcomes related to resource utilization, such as length of stay, can be predicted using ML methods in intensive care admissions[2]. In this pilot study, we aimed to evaluate the predictability of intensive care mortality and the need for intubation within the first 48 hours using ML methods based on routine data obtained at the time of admission to the intensive care unit.

Material and Methods

Study Design and Participants

This was a single-center, retrospective, observational pilot study. Adult patients aged 18 years or older who were admitted to the intensive care unit between December 1 and 31, 2025, were retrospectively screened. Patients with complete data for the prespecified admission variables and study outcomes were eligible for inclusion. Patients younger than 18 years and those with missing clinical, laboratory, or outcome data were excluded. Eligible patients were included consecutively according to the order of ICU admission.

Predictor Variables and Outcomes

Predictor variables were obtained at ICU admission or within the first few hours and included age, sex, primary diagnosis, APACHE II and SOFA scores, arterial blood gas parameters (pH, PaO_2, $PaCO_2$, and HCO_3^-), lactate, glucose, sodium, potassium, calcium, FiO_2, and the PaO_2/FiO_2 ratio. The two prespecified outcomes were ICU mortality and new endotracheal intubation within 48 hours after ICU admission. ICU mortality was evaluated in the entire cohort. The early-intubation outcome was evaluated only among patients who were not intubated at ICU admission. Patients already intubated at admission remained included in the mortality analysis but were excluded from the early-intubation analysis. Baseline intubation status was not used as a predictor in the early-intubation model.

Artificial Neural Network Development

The study included 10 unique patients. For ANN input construction, 13 prespecified clinical and laboratory features from each patient were restructured in long format, producing 130 feature-level records (10 patients×13 features). Each record included the corresponding patient-level demographic and diagnostic information, a coded clinical or laboratory feature, and the value of that feature; ICU mortality and intubation within 48 hours were specified as outcome labels. Accordingly, the 130 records represented feature-level entries derived from 10 patients and did not constitute 130 independent patients or independent clinical observations. This study employed a multilayer perceptron (MLP) architecture to train an ANN for the prediction of potentially detrimental clinical outcomes. We selected this feed-forward network architecture due to its capacity to characterize non-linear, intricate interactions between prevalent admission factors and life-threatening patient occurrences. The Levenberg-Marquardt algorithm, a powerful tool for improving MLP framework weights, was used to train the model.

The predictor inputs included age, sex, primary diagnosis, APACHE II and SOFA scores, arterial blood gas parameters, lactate, glucose, electrolytes, FiO_2, and the PaO_2/FiO_2 ratio. Baseline intubation status was included only in the ICU mortality model. For the early-intubation model, only patients who were not intubated at ICU admission were included, and baseline intubation status was excluded from the predictor set. Separate classification outputs were generated for ICU mortality and the need for new endotracheal intubation within 48 hours.

The 130 feature-level records were allocated to internal model-development subsets comprising 90 records for training, 20 for validation, and 20 for testing. This partitioning was used for exploratory internal model assessment and did not constitute independent patient-level validation. The MLP architecture included a single hidden layer with 20 neurons.

Model performance was evaluated using mean squared error (MSE), the correlation coefficient (R), margin of deviation (MoD), area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and F1-score. AUC was calculated from the continuous ANN output scores as the area under the receiver operating characteristic curve. Sensitivity, specificity, and F1-score were calculated from the corresponding binary classification results. These metrics were used for exploratory internal assessment of the feature-level training, validation, and testing subsets and were not considered measures of independent patient-level validation. The mathematical expressions for MSE, R, MoD, sensitivity, specificity, and F1-score are provided below.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (X_{targ(i)} - X_{pred(i)})^2 \quad (1)$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^{N} \left(X_{(i)} - X_{pred(i)}\right)^2}{\sum_{i=1}^{N} \left(X_{(i)}\right)^2}} \quad (2)$$

$$MoD\ (\%) = \left[\frac{X_{targ} - X_{pred}}{X_{targ}}\right] x\ 100 \quad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \quad (6)$$

In these equations, X_{targ} represents the actual clinical values, X_{pred} denotes the values predicted by the ANN model, while TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives obtained from the classification confusion matrix, respectively.

Ethical Statement

The protocol was approved by the Non-Interventional Clinical Research Ethics Committee of Niğde Ömer Halisdemir University Faculty of Medicine, Türkiye (Decision No: 2026/3, Date: 08.01.2026)

Results

Figure 1 presents the MSE values across 12 training epochs. At epoch 12, the MSE values were 5.46×10^{-15} for the training subset, 9.95×10^{-15} for the validation subset, and 5.63×10^{-15} for the testing subset. In the exploratory feature-level internal assessment, both the ICU mortality and early-intubation classifications yielded an AUC of 1.00, sensitivity of 100%, specificity of 100%, and an F1-score of 1.00.

Figure 1. The performance of the developed ANN model during its development phase

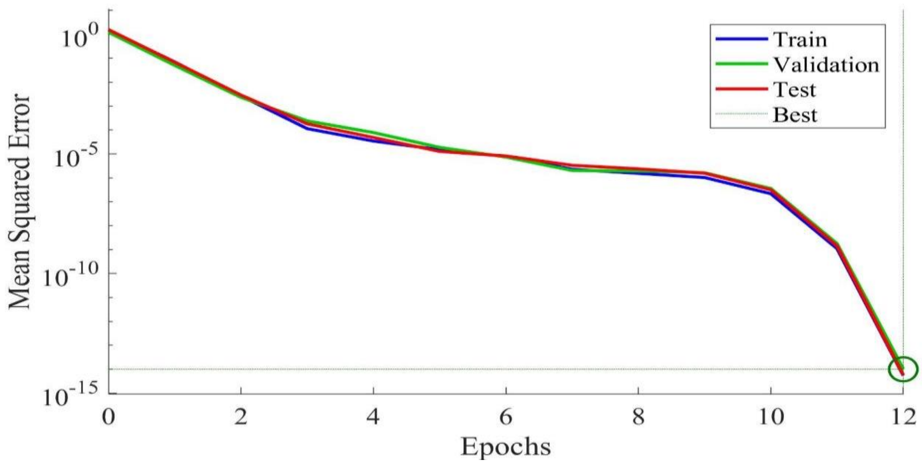
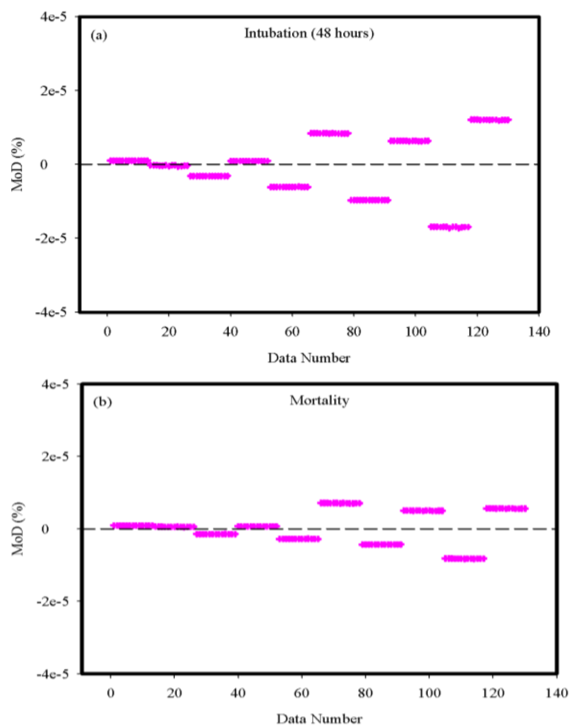


Figure 2 presents the MoD values across the feature-level records used for the internal assessment of each outcome. The deviations ranged approximately from -4×10^{-5} to $+4\times10^{-5}$. The mean MoD was -7.39×10^{-7}% for intubation within 48 hours and 3.32×10^{-7}% for ICU mortality.

Figure 2. Distribution of the MoD across the feature-level records used for the internal assessment of (a) intubation within 48 hours and (b) ICU mortality.



The R value was 0.99999 for the training, validation, and testing subsets.

Discussion

This pilot study provides preliminary evidence that an artificial neural network framework can be applied to routinely collected ICU admission data to explore mortality and the need for early intubation. However, considering the inclusion of only 10 patients and the high model complexity relative to the patient sample, the near-perfect internal performance estimates should be interpreted primarily as evidence of methodological feasibility rather than stable or generalizable predictive accuracy. These findings are exploratory and intended to guide the design of larger validation studies. Within this context, the overall direction of our findings is consistent with previous mortality prediction studies conducted using substantially larger datasets[1–6].

Among the predictor domains incorporated into the ANN framework, illness-severity scores, oxygenation measures, and metabolic variables have clear clinical relevance to the outcomes examined. APACHE II and SOFA scores reflect the extent of acute physiological disturbance and multisystem organ dysfunction and may therefore contain information relevant to both mortality and respiratory deterioration. Oxygenation variables, particularly FiO_2 and the PaO_2/FiO_2 ratio, reflect the severity of gas-exchange impairment and are clinically relevant to the potential need for early intubation. Similarly, elevated lactate and abnormalities in acid–base status, glucose, and electrolytes may indicate tissue hypoperfusion, metabolic stress, and loss of physiological homeostasis. However, because no separate formal feature-importance analysis was performed and the patient sample was very small, these observations should be regarded as clinically guided, hypothesis-generating interpretations rather than evidence of a definitive predictor ranking or independent prognostic effects. The clinical relevance of these predictor domains is consistent with previous machine learning studies of mortality and intubation outcomes in critically ill patients[7–13,19].

Early risk classification in patients with sepsis is challenging due to heterogeneous clinical courses. Methods addressing imbalanced data have been shown to improve sepsis mortality prediction, and machine learning models have also been developed for mortality prediction in patients with sepsis and in those with concomitant chronic kidney disease[10-12]. In this context, the hypothesis-generating value of pilot studies is important.

The risk of mortality and adverse clinical outcomes is substantially higher among critically ill patients requiring mechanical ventilation. Recent machine learning–based models have demonstrated the ability to predict hospital mortality in mechanically ventilated intensive care patients, highlighting the prognostic importance of early respiratory deterioration. Community- and cohort-based approaches predicting prolonged mechanical ventilation requirements further emphasize the clinical and resource-related implications of timely risk stratification. Studies analyzing factors associated with invasive mechanical ventilation using machine learning techniques, as well as models focusing on early mortality prediction in acute respiratory distress syndrome (ARDS), illustrate the evolving role of data-driven decision support systems in intensive care. A recent overview also underscores that artificial intelligence–based

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

applications may increasingly contribute to the management of mechanical ventilation in critical care practice[14–18].

One of the strengths of the present study is that all variables used in model development are readily available in routine clinical practice at the time of intensive care admission. Demonstrating that commonly encountered parameters, such as electrolyte imbalances and standard laboratory values, can meaningfully contribute to machine learning–based mortality prediction enhances the clinical applicability of such models. This observation is consistent with previous reports indicating that routinely measured biochemical variables may provide substantial predictive value when incorporated into machine learning frameworks[13].

Limitations

This study has several important limitations. First, the inclusion of only 10 patients substantially limits the statistical stability, representativeness, and generalizability of the findings. The high model complexity relative to the number of individual patients creates a substantial risk of overfitting and may contribute to the near-perfect internal performance estimates. Moreover, internal partitioning into training, validation, and testing subsets cannot substitute for validation in an independent patient cohort. In addition, the 130 feature-level records were derived from only 10 patients and were not statistically independent. Because partitioning was performed at the feature-record level, records originating from the same patient may have been represented in more than one model-development subset, creating a risk of information leakage and overly optimistic internal performance estimates. Therefore, the reported findings should be interpreted as exploratory evidence of methodological feasibility rather than as evidence of established clinical predictive accuracy. The retrospective, single-center design further limits external validity. Future studies should include substantially larger, prospectively collected, multicenter cohorts and perform independent external validation with patient-level separation across all model-development subsets.

Conclusions

This pilot study explored the methodological feasibility of applying an artificial neural network model trained on routinely collected ICU admission data to assess ICU mortality and the need for intubation within 48 hours. The developed multilayer perceptron model, which included 20 hidden neurons, yielded an R value of 0.99999 and very low mean squared error values across the internal training, validation, and testing subsets. However, given the inclusion of only 10 patients and the high model complexity relative to the patient sample, these near-perfect internal performance estimates may reflect overfitting and should not be interpreted as evidence of established clinical accuracy or generalizability. Routine admission variables may contain potentially informative signals for early risk assessment; however, this observation requires confirmation in substantially larger, prospective, multicenter cohorts with independent external validation. Therefore, the present findings should be considered preliminary and hypothesis-generating rather than evidence of a clinically applicable predictive model.

Conflict of Interest Statement: The authors declare no conflict of interest related to this study.

Financial Disclosure: The authors declare that no financial support was received during the conduct of this study.

Ethical Approval: Ethical approval for this study was obtained from the Non-Interventional Clinical Research Ethics Committee of Niğde Ömer Halisdemir University Faculty of Medicine (Decision No: 2026/3). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Author Contributions

SK: Conceptualization, study design, data collection, clinical interpretation, supervision, and writing. **ABÇ:** Writing - original draft, methodology, software, conceptualization, and formal analysis. **ŞBBP:** Conceptualization, study supervision, and critical review of the manuscript. **ZYT, MK:** Clinical patient care and general supervision. **SMB:** Data support

REFERENCES

1. Sun H, Kang M, Zhang H, Jia J, Wang Q. Machine learning for predicting mortality in intensive care unit patients: a prognostic performance systematic review and meta-analysis. *Nursing in Critical Care*. 2025;30(6):e70206.

2. Iwase S, Nakada T, Shimada T, et al. Prediction algorithm for ICU mortality and length of stay using machine learning. *Scientific Reports*. 2022;12:12912.

3. Pang K, Li L, Ouyang W, Liu X, Tang Y. Establishment of ICU mortality risk prediction models with machine learning algorithm using MIMIC-IV database. *Diagnostics*. 2022;12(5):1068.

4. Lim L, Gim U, Cho K, et al. Real-time machine learning model to predict short-term mortality in critically ill patients: development and international validation. *Critical Care*. 2024;28(1):76.

5. Thadani S, Wu TC, Wu DTY, et al. Machine learning-based prediction model for ICU mortality after continuous renal replacement therapy initiation in children. *Critical Care Explorations*. 2024;6(12):e1188.

6. Yang S, Sun Y, Wang T, et al. Machine learning-based prediction of mortality and multidrug-resistant infection risks in ICU patients with suspected infection: a prospective national multicenter cohort study. *BMC Infectious Diseases*. 2026;26:139.

7. Liu J, Duan X, Duan M, et al. Development and external validation of an interpretable machine learning model for the prediction of intubation in the intensive care unit. *Scientific Reports*. 2024;14:27174.

8. Li R, Xu Z, Xu J, et al. Predicting intubation for intensive care units patients: a deep learning approach to improve patient management. *International Journal of Medical Informatics*. 2024;186:105425.

9. Venturini M, Van Keilegom I, De Corte W, Vens C. Predicting time-to-intubation after critical care admission using machine learning and cured fraction information. *Artificial Intelligence in Medicine*. 2024;150:102817.

10. Amory IA, Rashidi Khazaee P, Yousefi S. Improving sepsis mortality prediction with machine learning using full region synthetic sampling approach. *Health Science Reports*. 2025;8(11):e71556.

11. Gao J, Lu Y, Ashrafi N, et al. Prediction of sepsis mortality in ICU patients using machine learning methods. *BMC Medical Informatics and Decision Making*. 2024;24:228.

12. An S, Ye Z, Che W, et al. Development and validation of machine learning models to predict in-hospital mortality in ICU patients with sepsis and chronic kidney disease. *BMC Infectious Diseases*. 2025;25(1):1504.

13. Silveira EC, Pretti SM, Santos BA, et al. Prediction of hospital mortality in intensive care unit patients from clinical and laboratory data: a machine learning approach. *World Journal of Critical Care Medicine*. 2022;11(5):317–329.

14. Li H, Ashrafi N, Kang C, Zhao G, Chen Y, Pishgar M. A machine learning-based prediction of hospital mortality in mechanically ventilated ICU patients. *PLoS ONE*. 2024;19(9):e0309383.

15. Weaver M, Goodin DA, Miller HA, et al. Prediction of prolonged mechanical ventilation in the intensive care unit via machine learning: a COVID-19 perspective. *Scientific Reports*. 2024;14:30173.

16. Zhang G, Xie Q, Wang C, Xu J, Liu G, Su C. Intelligent alert system for predicting invasive mechanical ventilation needs via noninvasive parameters: employing an integrated machine learning method with integration of multicenter databases. *Medical & Biological Engineering & Computing*. 2024;62(11):3445–3458.

17. Hannon DM, Syed JDA, McNicholas B, Madden M, Laffey JG. The development of a C5.0 machine learning model in a limited data set to predict early mortality in patients with ARDS undergoing an initial session of prone positioning. *Intensive Care Medicine Experimental*. 2024;12(1):103.

18. Misseri G, Piattoli M, Cuttone G, Gregoretti C, Bignami EG. Artificial intelligence for mechanical ventilation: a transformative shift in critical care. *Therapeutic Advances in Pulmonary and Critical Care Medicine*. 2024;19:29768675241298918.

19. El-Rashidy N, Tarek Z, Elshewey AM, et al. Multitask multilayer-prediction model for predicting mechanical ventilation and the associated mortality rate. *Neural Computing and Applications*. 2025;37:1321–1343.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 37-53.

Can Artificial Intelligence Reliably Detect Risky Eating Narratives? A Comparative Analysis of YouTube Extreme Diet Videos

İrem BAŞÖREN[*], Hilal AKSOY[**], Abdullah Furkan KORKMAZ[***], Zehra Margot ÇELİK[****]

Abstract

Aim: This descriptive study was conducted to identify characteristics that are related to binge eating in YouTube videos that have been shared with the keywords “extreme diet” and “diet challenge” and to compare classifications made by expert dietitians with those generated by an artificial intelligence model.

Method: A total of 49 YouTube videos that met the inclusion criteria were analyzed. The video characteristics were evaluated in relation to the view count, number of likes and comments, video duration, and time elapsed since upload**.** The quality of information provided in the video was assessed using the Modified DISCERN instrument, JAMA Benchmark Criteria, and Global Quality Score (GQS). A study-specific Binge Eating–Related Scoring System (BERS), consisting of 15 dichotomous items and based on video transcripts, was independently applied by both expert dietitians and the Gemini 3 Pro artificial intelligence model. In addition, 89,240 user comments were analyzed using R-based sentiment and emotion analysis techniques. Google Trends data for 2021–2026 were also evaluated. Statistical analyses were performed using SPSS 22.0.

Results: The median number of views was 2,043,167 (IQR: 1,043,285–5,550,269). AI-based BERS scores showed significant positive correlations with the number of views ($r=0.471$), likes ($r=0.453$), comments ($r=0.423$), and video length ($r=0.543$) ($p<0.05$). A moderate positive correlation was found between expert-rated and AI-generated BERS scores ($r=0.577$, $p<0.001$). Higher binge eating–related scores were negatively correlated with educational quality scores (GQS). Sentiment analysis revealed that 41.3% of comments were positive, and the most frequently identified emotions were trust and joy. Google Trends indicated a peak in “extreme diet” searches in 2026.

Conclusion: Highly engaging YouTube content often not only normalizes extreme eating behaviors but is also received positively by viewers. AI proved to be moderately reliable for identifying binge eating-related features, thus suggesting its potential as a scalable screening tool for monitoring harmful health trends. Interdisciplinary collaboration is required to ensure that AI-powered content analysis aligns with public health priorities.

Keywords: Nutrition, artificial intelligence, binge eating disorder, social media, YouTube.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 05.03.2026 **Kabul / Accepted:** 29.06.2026
DOI: https://doi.org/10.38079/igusabder.1903214
[*] Res. Asst., Istanbul Gelisim University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: ibasoren@gelisim.edu.tr ; MSc Student, Marmara University, Institute of Health Sciences, Istanbul, Türkiye. E-mail: basorenirem1@gmail.com ORCID https://orcid.org/0009-0007-1095-6603
[**] MSc Student, Hacettepe University, Institute of Informatics, Department of Health Informatics, Ankara, Türkiye. E-mail: hilalaksoy0674@gmail.com ORCID https://orcid.org/0009-0008-2715-6597
[***] BSc Student, Gazi University, Faculty of Science, Department of Mathematics, Ankara, Türkiye. E-mail: afkorkmaz@outlook.com ORCID https://orcid.org/0009-0006-6485-0567
[****] (Corresponding Author) Asst. Prof., Marmara University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: zcelik@marmara.edu.tr ORCID https://orcid.org/0000-0002-2138-5996

Yapay Zekânın Riskli Yeme Anlatılarını Belirlemedeki Güvenilirliği: YouTube Radikal Diyet Videolarının Karşılaştırmalı Analizi

Öz

Amaç: Bu tanımlayıcı çalışma, “extreme diet” ve “diet challenge” anahtar kelimeleri ile paylaşılan YouTube videolarında tıkınırcasına yeme ile ilişkili özelliklerin varlığını değerlendirmek ve uzman diyetisyenler tarafından yapılan sınıflandırmalar ile bir yapay zekâ modeli tarafından oluşturulan sınıflandırmaları karşılaştırmak amacıyla yapılmıştır.

Yöntem: Dahil edilme kriterlerini karşılayan toplam 49 YouTube videosu analiz edilmiştir. Videoların görüntülenme sayısı, beğeni sayısı, yorum sayısı, süresi ve yüklenme tarihinden itibaren geçen süre kaydedilmiştir. Bilgi kalitesi Modifiye DISCERN aracı, JAMA (Journal of the American Medical Association) kriterleri ve Küresel Kalite Skoru (GQS) kullanılarak değerlendirilmiştir. Çalışmaya özgü 15 maddelik Tıkınırcasına Yeme ile İlişkili Puanlama Sistemi (BERS), uzman değerlendiriciler ve Gemini 3 Pro yapay zekâ modeli tarafından (video transkriptleri üzerinden) bağımsız olarak uygulanmıştır. Ayrıca 89.240 kullanıcı yorumu R programlama dili kullanılarak duygu ve duygu durumu analizi yöntemleriyle incelenmiştir. 2021–2026 yıllarına ait Google Trends verileri de değerlendirilmiştir. İstatistiksel analizler SPSS 22.0 programı ile gerçekleştirilmiştir.

Bulgular: Medyan görüntülenme sayısı 2.043.167 (IQR: 1.043.285–5.550.269) olarak bulunmuştur. Yapay zekâ temelli BERS puanları ile görüntülenme ($r=0,471$), beğeni ($r=0,453$), yorum sayısı ($r=0,423$) ve video süresi ($r=0,543$) arasında istatistiksel olarak anlamlı pozitif ilişki saptanmıştır ($p<0,05$). Uzman ve yapay zekâ BERS puanları arasında orta düzeyde pozitif korelasyon belirlenmiştir ($r=0,577$, $p<0,001$). Tıkınırcasına yeme ile ilişkili daha yüksek puanların eğitimsel kalite (GQS) ile negatif ilişkili olduğu görülmüştür. Duygu analizinde yorumların %41,3’ünün pozitif olduğu; en sık belirlenen duyguların güven ve sevinç olduğu saptanmıştır. Google Trends verileri, 2026 yılında “extreme diet” aramalarında zirveye ulaşıldığını göstermiştir.

Sonuç: Yüksek etkileşim alan YouTube içerikleri, aşırı yeme davranışlarını normalleştirebilmekte ve izleyiciler tarafından olumlu algılanabilmektedir. Yapay zekâ sistemi, tıkınırcasına yeme ile ilişkili özelliklerin belirlenmesinde orta düzeyde güvenilirlik göstermiş olup zararlı sağlık anlatılarının izlenmesinde ölçeklenebilir bir tarama aracı olarak potansiyel sunmaktadır. Yapay zekâ destekli içerik analizinin halk sağlığı öncelikleri ile uyumlu hale getirilmesi için disiplinler arası işbirliği gereklidir.

Anahtar Sözcükler: Beslenme, yapay zekâ, tıkınırcasına yeme bozukluğu, sosyal medya, YouTube.

Introduction

In recent years, there has been a rapid rise in the use of digital resources to meet a wide range of needs; these include seeking advice, entertainment, cooking, and coping with daily stress[1]. For many individuals, the internet, particularly social media platforms, has taken on the role of being the primary source for health-related guidance, lifestyle information, and nutrition advice[2]. YouTube is one of the most widely used platforms, with over two billion active users globally. It has become increasingly influential in disseminating advice, daily routines, and personal experiences through video blogs (vlogs) and user-generated content[2,3]. Video-sharing platforms such as YouTube facilitate seamless access to fitness, dieting, body image, and lifestyle-related content. Young adults frequently state that they use YouTube for health and nutrition advice, referencing its instructional and relatable format[4]. Although social media has been used in nutrition and health promotion interventions, there is growing evidence that suggests frequent exposure to diet- and body-focused content can be associated with adverse outcomes[5,6]. A systematic review indicates that particularly among adolescents and young adults, there is a significant association between social media use and eating disorders, including restrictive eating and binge eating[7]. Consequently, while social media can be used to promote healthy behavior, repeated exposure to extreme dietary practices may normalize harmful eating patterns[7,8].

Binge eating disorder (BED) is characterized by recurrent episodes of consuming large amounts of food within a short period; this is accompanied by loss of control and psychological distress, while there is no compensatory behavior[9]. Extreme dieting, which is also considered to be an eating disorder, involves severe restriction and rigid control of food intake, or prolonged fasting, and most often is pursued to ensure rapid weight loss or body modification. Such practices are associated with significant psychological and physiological risks and have been closely associated with body dissatisfaction or maladaptive weight-control motivations[10]. It is important to note that YouTube diet videos frequently frame extreme practices, such as prolonged fasting, not just as health strategies, but also as morally commendable or disciplined behavior, thus potentially legitimizing eating patterns that are highly restrictive[11].

It is thought that exposure to video content related to eating that is found on platforms such as YouTube can influence the eating attitudes and behavior of the audience[12]. Studies that examine content that is focused on eating, including videos that portray excessive consumption or restrictive practices, indicate that repeated exposure may not only normalize problematic behaviors but also shape perceptions of what is acceptable conduct regarding diet[13,14]. These findings raise concerns that diet-related video trends may contribute to the development of unhealthy eating behaviors, particularly among young and vulnerable audiences[1,12].

Diet-related content on YouTube is widespread; as a result it is important not only to examine the messages conveyed in these videos, but also to evaluate the tools that have been used to analyze the content. Although the use of artificial intelligence (AI) in health-related research has increased, there are still questions regarding how reliable AI-based analyses are and how much they align with expert human evaluation. Previous studies have reported substantial but imperfect concordance between AI classification systems and human raters, highlighting the need for further validation before there can be a large-scale application of AI systems in this area.

Although previous research has evaluated the quality of YouTube content related to specific eating disorders, such as avoidant/restrictive food intake disorder (ARFID), these studies have relied primarily on human assessment[11,15]. There is little evidence whether AI-based tools can identify risky or extreme diet-related narratives as reliably as human-based judgment.

In this study, the aim was to evaluate themes related to eating disorders in YouTube videos that have been shared under the keywords “extreme diet” and “diet challenge,”

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

and to compare classifications made by a registered dietitian and that of the Gemini 3 Pro artificial intelligence model using a structured binary scoring system. A secondary objective was to assess the agreement and potential applicability of AI-based tools for identifying risky eating-related content on social media platforms.

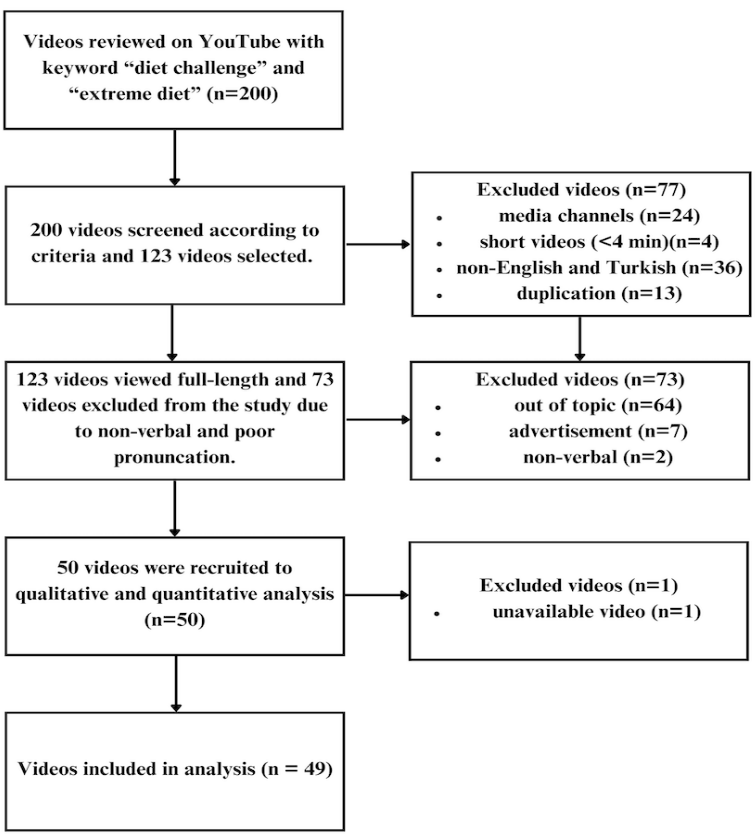
Material and Methods

Search strategy

Video selection was conducted independently by two dietitian researchers on YouTube in October 2025 using newly created accounts to reduce algorithmic and user-related biases. Searches were performed using the keywords "extreme diet" and "diet challenge." Videos published within the past five years, presented in English or Turkish, lasting at least four minutes, and uploaded by individual content creators were included. Promotional, duplicated, irrelevant, or non-eligible videos were excluded.

For each keyword, the first five pages of search results were screened, as users typically interact with content appearing on initial result pages (20 videos per page×5 pages=100 videos)[16,17]. Videos were sorted by view count, resulting in an initial pool of 200 videos. Of these, 50 met the inclusion criteria; however, one video later became unavailable, leaving a final sample of 49 videos. The selection process is illustrated in Figure 1.

Figure 1. Flowchart of the video selection process



Video characteristics

In evaluating the videos, the number of views, likes, and number of comments, video length (seconds), time since upload days and comment content were taken into account. These variables were included in the analysis to reflect the level of video engagement and user interaction.

Reliability, validity, and quality assessment

The Modified DISCERN (mDISCERN) instrument was used to evaluate the quality and comprehensiveness of the videos. This tool includes five criteria assessing whether the video is clear and understandable, based on reliable sources, presents balanced information, provides additional references, and addresses areas of uncertainty. Each item is scored dichotomously (1=yes, 0=no), yielding a total score ranging from 0 to 5, where higher scores indicate better quality and comprehensiveness of the content[18].

The Journal of the American Medical Association (JAMA) Benchmark Criteria were used to assess the reliability and quality of the medical information based on four components: authorship, attribution, disclosure, and currency. Each criterion received one point, resulting in a total score of between 0 and 4. Scores of 0–1 indicate insufficient information, 2–3 partially sufficient information, and 4 high-quality information[19].

To evaluate the overall educational quality of the videos, the Global Quality Score (GQS) was used. With this scale the clarity, quality, and usefulness of information was assessed using a five-point Likert scale, where 1 represents poor quality and minimal educational value and 5 represents excellent quality with high educational value[20].

All videos were independently evaluated by two researchers using these three assessment tools. Discrepancies between raters were resolved through discussion and consensus.

Sentiment–emotion analysis of videos with R programming language

Sentiment analysis of comments from 49 YouTube videos related to “diet challenges” and “extreme diets” was conducted using the R 4.5.0 statistical programming environment. User comments stored in JSON Lines (.jsonl) format were analyzed using natural language processing (NLP) techniques. JSONL files were imported using the jsonlite package[21], merged into a single dataset, and only the text variable containing comment content was included in the analysis.

Text tokenization and sentiment classification were performed using the tidytext package[22]. Comments were tokenized with the unnest_tokens() function, and only alphabetic words consisting of at least three characters were retained. Text cleaning and data manipulation were conducted using the dplyr and stringr packages[23,24].

A two-stage stop-word filtering procedure was applied. First, the standard English stop-words from the list in tidytext were removed[18]. Second, commonly used but context-independent YouTube comment terms (e.g., “video,” “watch,” “day,” “people,” “bro,” “gonna,” and “ate”) were excluded using a manually created extended stop-word list. After preprocessing, a total of 89,240 comments were included in the analysis.

Sentiment classification was conducted using the AFINN lexicon, which categorizes text as positive, negative, or neutral. Emotional categories such as anger, sadness, fear, and joy were identified using the NRC lexicon. Visualization of results was performed with the ggplot2 package[25]. The most frequent words were presented using two visualization approaches: a bar plot displaying the ten most frequent words and a word cloud generated from word frequencies. In the word cloud, only words above a minimum frequency threshold were included, and word size was scaled proportionally to frequency[26]. The maximum number of displayed words was limited and the Dark2 palette from RColorBrewer was applied to improve visual clarity[27].

All text mining, sentiment analysis, and visualization procedures were conducted in R using the *jsonlite, dplyr, stringr, tidytext, ggplot2, wordcloud*, and *RColorBrewe*r packages[21-27].

Study-specific binge eating–related scoring system (BERS)

Since mDISCERN, GQS, and JAMA benchmark criteria do not specifically evaluate binge eating–related characteristics in social media videos, a study-specific scoring system was developed to assess excessive consumption framing, behavioral modeling, triggering narratives, and health-related messaging in videos identified using the keywords “extreme diet” and “diet challenge.” The BERS framework was designed to identify content features that may normalize or encourage maladaptive eating behaviors rather than to diagnose eating disorders. Development of the scoring system was informed by a targeted review of literature on binge eating–related behavioral indicators[7,28] and adapted from previously published topic-specific content analysis approaches used in video-based research[29] (Table 1).

The checklist consisted of 15 criteria, each scored dichotomously (1=present, 0=absent) based on verbal or visual presentation within the video. Total scores were analyzed using K-means clustering to classify videos into three groups representing lower, moderate, and higher levels of binge eating–provoking content.

Prior to the main analysis, a pilot assessment was conducted in which three researchers independently scored three videos, while the same videos were evaluated by the Gemini AI tool using identical scoring criteria. Partial agreement between human and AI scores supported proceeding with the main analysis.

All videos were independently evaluated in the main study by three researchers who used the BERS checklist. To avoid any possible bias, separate AI-based scoring was carried out by a fourth researcher who used Gemini to evaluate only the transcripts of the videos.

A structured Python-based workflow was used to score the video transcription and AI scoring. Audio streams were extracted with yt-dlp and processed by FFmpeg. Faster-Whisper (Large-V3) model was used for automatic speech recognition; predefined parameters (beam size=5, float16 precision, temperature schedule 0–0.4, and voice activity detection enabled) were used to accomplish this. Transcripts were stored as plain text and time-stamped JSON segment files, while a manifest system was used to record processing steps and outputs.

BERS scoring was carried out with the Gemini 3 Pro Preview model using standardized instructions. Each transcript was evaluated independently, and all 15 BERS items were scored dichotomously (0/1). Model parameters were fixed (temperature=0.1, top_p=0.95, thinking level=high), while structured JSON outputs were enforced so that there would be consistent formatting across evaluations.

Thus, each video received both researcher-derived and AI-generated BERS scores; the relationship between these scores was examined during statistical analysis.

Table 1. Binge Eating–Related Scoring System (BERS)

1	Does the video present excessive consumption as an achievement, performance, or challenge?
2	Are extremely high calorie amounts emphasized in an enthusiastic or celebratory manner?
3	Is excessive eating framed as easy, repeatable, or normalized within the video narrative?
4	Is the eating episode described using highly sensory or exaggerated praise?
5	Is loss of control over eating explicitly expressed?
6	Are contextual triggers related to eating behavior presented?
7	Is reward-based eating language used?
8	Is the rapid consumption of large quantities emphasized?
9	Is the continuation of eating despite physical discomfort described?
10	Are feelings of shame, guilt, or regret explicitly expressed?
11	Are compensatory physical activities presented as a response to consumption?
12	Are compensatory behaviors related to eating implied or stated?
13	Does the video include a direct call for imitation by the audience?
14	Is escalation or intent to increase the intensity of consumption presented?
15	Is body weight or body-related outcome explicitly emphasized?

Questions were answered as yes or no, yes = 1-point, no = 0-point

Ethical Statement

Ethical considerations were observed throughout the study. User anonymity was maintained and no personally identifiable information was collected or reported. All analyses were conducted in accordance with the relevant ethical guidelines for the use of user-generated online content.

Statistical Analysis

Statistical analysis was performed using SPSS software (version 22, SPSS Inc., Chicago, IL, USA). The Shapiro-Wilk test was used to establish distributional properties of the variables. As none of the continuous variables demonstrated a normal distribution, the data were expressed as the median and interquartile range (IQR). Spearman's rank correlation analysis was used to examine relationships between quantitative variables. K-means clustering analysis was applied to identify distinct groups within the dataset. A p-value of less than 0.05 was considered statistically significant.

Results

The general characteristics of the evaluated videos are presented in Table 2. According to the table, the median values (interquartile range (IQR)) for total views, likes, and comments were 2,043,167 (1,043,285–5,550,269), 68,441 (27,120–125,465) and 3,014 (1,446–6,136) respectively. The median video length was 1,099 seconds (858–1,491), and the median time since upload was 1,371 days (650–1,799).

Table 2. General characteristics of videos

	Median (IQR)	**Mean ± S.D.**
Total Number of Views	2 043 167 (1 043 285–5 550 269)	5 202 917.02 ± 8359237.84
Total Number of Likes	68 441 (27 120–125 465)	125 895.76 ± 190742.27
Total Number of Comments	3 014 (1 446–6 136)	7 731.43 ± 15303.3
Video Length (seconds)	1 099 (858–1 491)	1 282.67 ± 714.78
Time Since Upload (days)	1 371 (650–1 799)	1 211.22 ± 686.58

IQR: interquartile range

Table 3 presents correlations between video characteristics, sentiment scores, and the mDISCERN, JAMA, GQS, BERS, and BERS-AI scoring systems. In the BERS system, a weak positive correlation was observed between video length and BERS score ($r=0.323$, $p=0.024$). In the BERS-AI system, moderate positive correlations were found with the number of views ($r=0.471$, $p=0.001$), likes ($r=0.453$, $p=0.001$), comments ($r=0.423$, $p=0.002$), and video length ($r=0.543$, $p<0.001$). Time since upload showed a significant negative correlation with BERS-AI scores ($r=-0.355$, $p= 0.012$). BERS-AI scores were also positively correlated with positive (r=0.325, p=0.022), negative (r=0.342, p=0.016), and neutral (r=0.437, p=0.002) sentiment categories.

Table 3. Correlations between video features and sentiment scores with the mDISCERN, JAMA and GQS scoring systems.

	mDISCERN		**JAMA**		**GQS**		**BERS**		**BERS-AI**	
	r	**p**	**r**	**p**	**r**	**p**	**r**	**p**	**r**	**p**
Features of Videos										
Views	-0.093	0.525	0.056	0.701	-0.050	0.733	0.247	0.087	.471**	0.001
Likes	-0.057	0.697	0.008	0.954	-0.048	0.742	0.201	0.165	.453**	0.001
Number of Comments	-0.139	0.340	-0.043	0.767	-0.163	0.263	0.171	0.239	.423**	0.002
Video Length	0.149	0.307	0.003	0.986	0.082	0.576	.323*	0.024	.543**	0.000
Time Since Upload	0.137	0.349	0.088	0.547	0.207	0.153	-0.130	0.373	.355*	0.012
Sentiments & Emotions										
Positive	-0.033	0.824	-0.030	0.840	-0.010	0.946	0.033	0.820	.325*	0.022
Negative	0.051	0.729	-0.064	0.664	-0.030	0.837	0.069	0.640	.342*	0.016
Neutral	-0.166	0.256	-0.175	0.230	-0.244	0.090	0.203	0.162	.437**	0.002

r, Spearman's rho; *p < 0.05, **p < 0.01

mDISCERN: Modified DISCERN; JAMA: Journal of the American Medical Association; GQS: Global Quality Score; BERS: Binge Eating–Related Scoring System; BERS-AI: Binge Eating–Related Scoring System- Artificial Intelligence

The correlations between the DISCERN, GQS, BERS, and BERS-AI scoring systems are presented in Table 4. A statistically significant, moderate, positive correlation was observed between DISCERN and GQS scores (r=0.641, p<0.001). GQS scores showed a weak negative correlation with BERS scores (r=-0.297, p=0.038) and a moderate negative correlation with BERS-AI scores (r=-0.370, p=0.009). A statistically significant moderate positive correlation was identified between BERS and BERS-AI scores (r=0.577, p<0.001).

Table 4. Correlations between scores

	DISCERN		JAMA		GQS		BERS		BERS-AI	
	r	**p**	**r**	**p**	**r**	**p**	**r**	**p**	**r**	**p**
DISCERN	1.000	-								
JAMA	0.202	0.163	1.000	-						
GQS	.641**	0.000	0.208	0.151	1.000	-				
BERS	-0.086	0.555	-0.060	0.680	-.297*	0.038	1.000	-		
BERS-AI	-0.160	0.272	0.043	0.769	-.370**	0.009	.577**	0.000	1.000	-

r, Spearman's rho; *$p < 0.05$, **$p < 0.01$

mDISCERN: Modified DISCERN; JAMA: Journal of the American Medical Association; GQS: Global Quality Score; BERS: Binge Eating–Related Scoring System; BERS-AI: Binge Eating–Related Scoring System- Artificial Intelligence

Figure 2 shows the distribution of the overall sentiment categories. Of all the comments analyzed, 41.3% (n=36,831) were classified as positive, 37.7% (n=33,629) as neutral, and 21.0% (n=18,780) as negative.

Emotion analysis based on the NRC lexicon revealed that the most frequently identified emotions were trust (11.78%, n=43,409) and joy (11.32%, n=41,742). This were followed by anticipation (9.52%), fear (8.21%), sadness (6.82%), anger (6.37%), disgust (5.96%), and surprise (5.15%).

Figure 2. Percentages of sentiments and emotions expressed in comments on videos

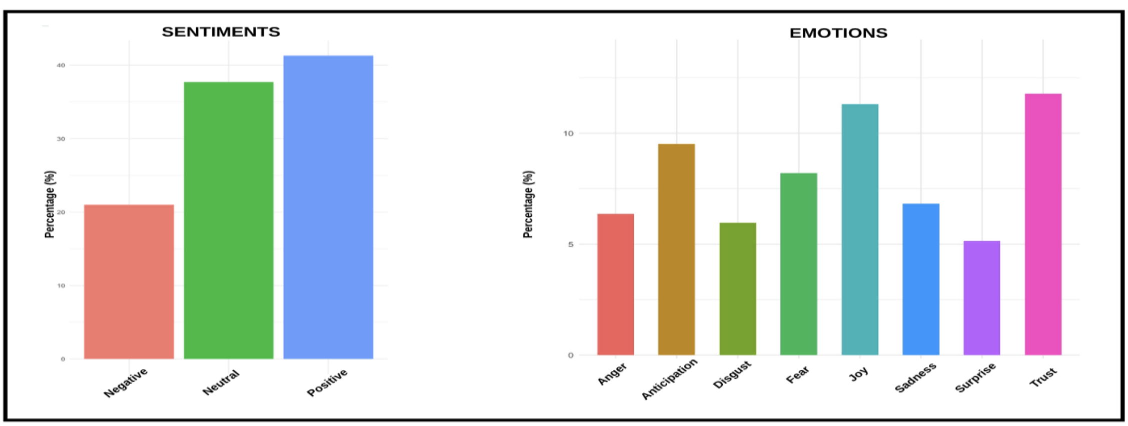


Figure 3 shows the most frequently used words in user comments. The most commonly used term was "eat" (n=8 329), followed by "food" (n=7 883), "diet" (n=5 526), and "love" (n=5 362). Other frequently used words included "eating" (n=4 832), "challenge" (n=3 481), "calories" (n=3 107), "weight" (n=3 048), "water" (n=3 011), and "feel" (n=2 315).

The word cloud visualisation (Figure 3) further demonstrates that terms such as “food”, “eat”, “diet,” “eating”, “challenge,” and “love” appeared prominently in the comment corpus.

Figure 3. The number of times words were used in the comments, and the Word cloud showing the most frequently used words in the comments

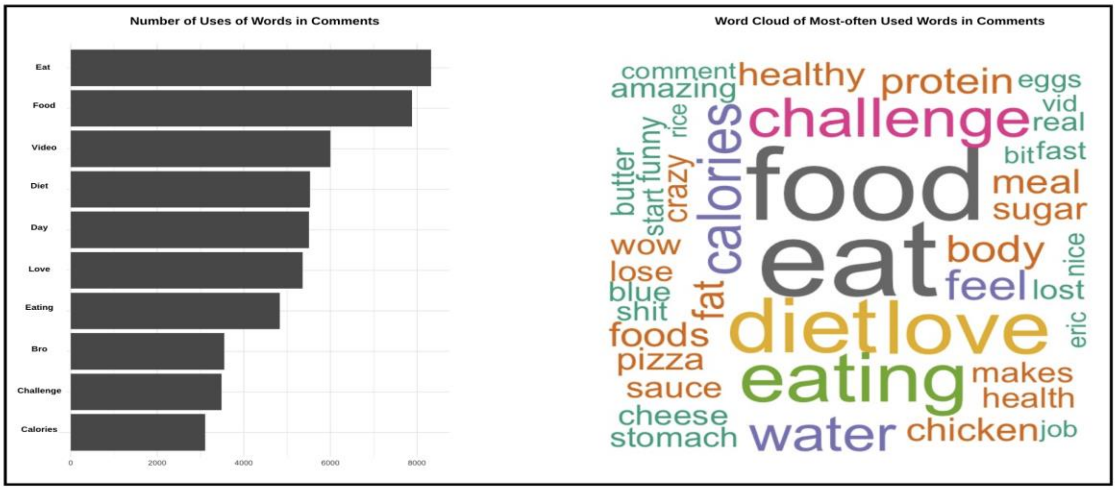
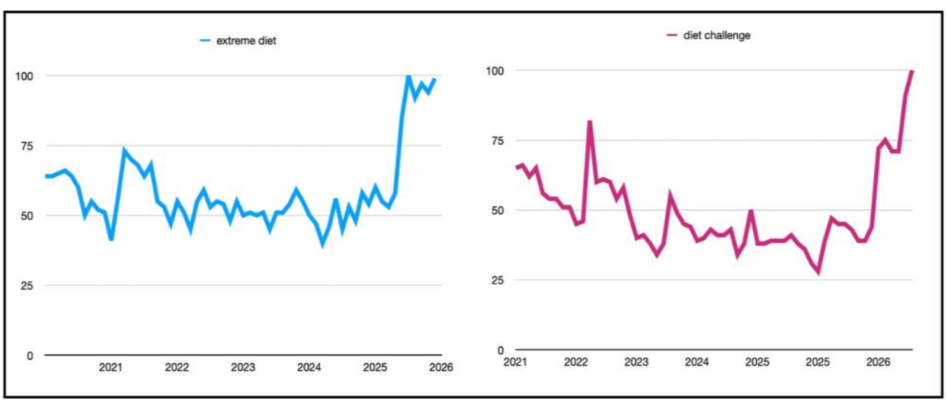


Figure 4 presents Google Trends data for the keywords “extreme diet” and “diet challenge” between 2021 and 2026. Search interest in the term “extreme diet” fluctuated between approximately 40 and 70 index points between 2021 and 2024. A marked increase was observed beginning in late 2025, with search interest reaching peak values close to 100 in 2026.

Similarly, search interest in “diet challenge” varied between approximately 30 and 70 index points from 2021 to 2024. A significant increase in search interest was observed towards the end of 2025, with the search index approaching 100 in 2026.

Figure 4. Increase in search keywords “extreme diet” and “diet challenge” on YouTube (https://trends.google.com)



Discussion

This study evaluated YouTube videos that were shared under the keywords "extreme diet" and "diet challenge" by combining content quality measures, BERS, BERS-AI classification, and sentiment analysis. Overall, the findings suggest that it is possible that highly engaging diet-related content can normalize extreme eating behaviors. At the same time, AI-based systems demonstrate potential for large-scale monitoring although there is still a need for validation and human oversight[30,31].

Higher BERS-AI scores were associated with greater engagement (views, likes, comments, and video length), indicating that videos which have stronger characteristics related to binge eating tend to reach wider audiences. Previous research similarly shows that performative eating and exaggerated consumption are amplified by platform dynamics which prioritize viewing time and user engagement[13].

The fact that there is a negative correlation between time since upload and BERS-AI scores suggests that more recent videos have more characteristics related to binge eating. This finding is consistent with studies which link increased exposure to dieting and appearance-focused social media content with greater body dissatisfaction and the risk of developing eating disorders[10,32].

A moderate positive correlation was observed between DISCERN and GQS scores, indicating consistency between the educational quality indicators. However, there was a negative association for both BERS and BERS-AI with GQS; this suggests that videos which have a higher educational quality are less likely to include characteristics that might provoke binge eating. Similar findings have been reported in studies which demonstrate that evidence-based information does not necessarily have a correlation with YouTube popularity[13,33]. Moreover, the absence of a correlation between JAMA scores and binge eating–related scores indicates that structural transparency alone is insufficient to prevent the normalization of extreme eating behaviors[34].

Sentiment analysis showed that the majority of comments were positive or neutral, with trust and joy being the most frequently identified emotions. BERS-AI scores were positively associated with positive, negative, and neutral sentiments, thus suggesting that videos which have more characteristics related to binge eating create greater emotional engagement. Similar patterns have been observed in studies on mukbang and diet-related content, where problematic eating representations coexist with strong viewer approval[35,36]. Experimental research has also demonstrated that exposure to diet-focused or idealized body content can increase body dissatisfaction and unhealthy eating habits, particularly among young audiences[37,38].

The comparison between BERS and BERS-AI scores revealed a moderate positive correlation, indicating partial agreement between expert evaluation and AI-based classification. This finding is in keeping with previous research, demonstrating that AI systems are able to approximate human judgments in health-related content analysis but are unable to fully replace expert evaluation[39]. The differences most likely arise due to

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

the visual and contextual cues that experts work from, whereas BERS-AI relies only on transcripts. Visual elements and editing styles may therefore influence behavioral modeling in ways that are not perceived through text-based analysis. These findings support the use of AI as a preliminary screening tool within a human-in-the-loop framework[34].

Overall, the results highlight a tension between platform engagement mechanisms and public health priorities. Videos that have more characteristics related to binge eating had higher engagement and produced more positive emotional responses; at the same time, educational quality indicators were inversely related to behavioral risk features. These findings reinforce concerns that social media popularity does not reliably reflect informational quality[30] and that repeated exposure to diet-focused content may increase the risk of developing eating disorders among young audiences[37]. Importantly, the study does not suggest that all diet challenges or extreme diet content are inherently harmful, but rather identifies narrative features which may increase vulnerability when repeatedly reinforced.

Strengths and Limitations

This study seeks to integrate validated quality assessment tools, a theory-driven risk scoring system (BERS), AI-based classification (BERS-AI), and sentiment analysis. The comparison of expert ratings and AI outputs makes a significant contribution to research into the automated evaluation of digital health content. However, there are several limitations that should be noted. The sample size was relatively small and restricted to English and Turkish videos. Any causal interpretation is prevented by the cross-sectional design. The AI analysis relied only on transcripts and did not include visual or multimodal features, and only one AI model was applied. In addition, contextual nuances may not be captured by lexicon-based sentiment analysis.

Future research should include larger samples, use multimodal AI approaches, and adopt longitudinal designs to provide a better understanding of the potential behavioral impact of extreme diet-related content.

Conclusion

It is possible that engaging YouTube videos of extreme dieting and diet challenges can normalize maladaptive eating behaviors. The association between higher BERS-AI scores and engagement metrics suggests that algorithmic visibility may amplify content depicting greater behavioral risk. Artificial intelligence had moderate agreement with human expert evaluation, thus indicating a potential for AI as a scalable screening tool to identify risky eating-related narratives; however, it is important to note that this application should complement—not replace—expert judgment. As digital platforms increasingly shape dietary norms, collaboration between nutrition science, public health, and AI is essential in order to create and support responsible health communication.

REFERENCES

1. Kircaburun K, Harris A, Calado F, Griffiths MD. The psychology of mukbang watching: A scoping review of the academic and non-academic literature. *Int J Ment Health Addict*. 2020;19(4):1190-1213.

2. Kiss A, Tompa O, Lakner Z, Unger-Plasek B, Temesi Á, Soós S. Conveying sport nutrition information in YouTube videos: A qualitative content analysis of dietary advice and ways of communication. *Curr Dev Nutr*. 2025;9(9):107525.

3. Kiss A, Soós S, Temesi Á, Unger-Plasek B, Lakner Z, Tompa O. Evaluation of the reliability and educational quality of YouTube videos on sport nutrition topics. *J Int Soc Sports Nutr*. 2023;20(1):2278632.

4. Lim MSC, Molenaar A, Brennan L, Reid M, McCaffrey T. Young adults' use of different social media platforms for health information: Insights from web-based conversations. *J Med Internet Res*. 2022;24(1):e23656.

5. Sanzari CM, Gorrell S, Anderson L, et al. The impact of social media use on body image and disordered eating behaviors: Content matters more than duration of exposure. *Eat Behav*. 2023;49:101722.

6. Rounsefell K, Gibson S, McLean S, et al. Social media, body image and food choices in healthy young adults: A mixed methods systematic review. *Nutr Diet*. 2020;77(1):19-40.

7. Moffitt L, Thomas MS. Influence of social media on dietary habits and nutritional choices in adolescents: A systematic review of the evidence. *Int J Nutr Food Sci*. 2025;14(4):211-225.

8. Liu L, Yang J, Tan F, Luo H, Chen Y, Zhao X. Web-based video platforms as sources of information on body image dissatisfaction in adolescents: Content and quality analysis of a cross-sectional study. *JMIR Form Res*. 2025;9:e71652.

9. Giel KE, Bulik CM, Fernandez-Aranda F, et al. Binge eating disorder. *Nat Rev Dis Primers*. 2022;8(1):16.

10. Dane A, Bhatia K. The social media diet: A scoping review to investigate the association between social media, body image and eating disorders amongst young people. *PLoS Glob Public Health*. 2023;3(3):e0001091.

11. Chiu MC, Penders B. “Maybe a long fast is good for you”: Health conceptualisations in YouTube diet videos. *Front Commun*. 2021;6:625906.

12. Wu Y, Kemps E, Prichard I. Mind your figure! Watch, but don’t eat: A content analysis of eating and appearance-related messages in eating videos on social media. *Body Image*. 2025;52:101883.

13. Akçay E, Bahadır B. Are YouTube videos a useful source of information on avoidant/ restrictive food intake disorder? *Arch Curr Med Res*. 2025;6(1):51-60.

14. Ferrey A, Ashworth G, Cabling M, Rundblad G, Ismail K. A thematic analysis of YouTube comments on a television documentary titled 'Diabulimia: The World's most dangerous eating disorder'. *Diabet Med*. 2023;40(5):e15025.

15. Suresh A, Pallempati LL, Saxena P, Ansari A, Bassi R, Bhandari A. Exploring YouTube videos about anorexia nervosa on the basis of reliability, popularity, and contributions of healthcare professionals: A cross-sectional study. *Cureus*. 2023;15(11):e48095.

16. Gimenez-Perez G, Robert-Vila N, Tomé-Guerreiro M, Castells I, Mauricio D. Are YouTube videos useful for patient self-education in type 2 diabetes? *Health Informatics J*. 2020;26(1):45-55.

17. Megaly M, Khalil C, Tadros B, Tawadros M. Evaluation of educational value of YouTube videos for patients with coeliac disease. *Int J Celiac Dis*. 2016;4(3):102-104.

18. Du RC, Zhang Y, Wang MH, Lu NH, Hu Y. TikTok and Bilibili as sources of information on Helicobacter pylori in China: A content and quality analysis. *Helicobacter*. 2023;28(5):e13007.

19. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveat lector et viewor—Let the reader and viewer beware. *JAMA*. 1997;277(15):1244-1245.

20. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Van Zanten SV. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. *Am J Gastroenterol*. 2007;102(9):2070-2077.

21. Ooms JJ. A simple and robust JSON parser and generator for R. CRAN website. https://cran.r-project.org/web/packages/jsonlite/. Published 2022. Accessed January 28, 2026.

22. Silge J, Robinson D. tidytext: Text mining using tidy data principles. CRAN website. https://cran.r-project.org/web/packages/tidytext/. Published 2022. Accessed December 4, 2022.

23. Wickham H, François R, Henry L, Müller K. dplyr: A grammar of data manipulation. CRAN website. https://cran.r-project.org/web/packages/dplyr/. Published 2022. Accessed December 4, 2022.

24. Wickham H. stringr: Simple, consistent wrappers for common string operations. CRAN website. https://cran.r-project.org/web/packages/stringr/. Published 2022. Accessed December 4, 2022.

25. Wickham H, Winston C, Henry L, et al. ggplot2: Create elegant data visualisations using the grammar of graphics. CRAN website. https://rdrr.io/cran/ggplot2/. Published 2022. Accessed January 28, 2026.

26. Fellows I. wordcloud: Word clouds. CRAN website. https://cran.r-project.org/web/packages/wordcloud/. Published 2018. Accessed December 4, 2022.

27. Neuwirth E. RColorBrewer: ColorBrewer palettes. CRAN website. https://cran.r-project.org/web/packages/RColorBrewer/. Published 2022. Accessed December 4, 2022.

28. Baykan GB, Batıgün AD. Reliability and validity of the Binge Eating Scale-Turkish form. *Turk J Clin Psychiatry*. 2024;27:230-238.

29. Ferhatoglu MF, Kartal A, Ekici U, Gurkan A. Evaluation of the reliability, utility, and quality of the information in sleeve gastrectomy videos shared on open access video sharing platform YouTube. *Obes Surg*. 2019;29(5):1477-1484.

30. Tan D, Huang Y, Liu M, Li Z, Wu X, Huang C. Identification of online health information using large pretrained language models: Mixed methods study. *J Med Internet Res*. 2025;27:e70733.

31. Benítez-Andrades JA, Alija-Pérez JM, Vidal ME, Pastor-Vargas R, García-Ordás MT. Traditional machine learning models and BERT-based automatic classification of tweets about eating disorders: Algorithm development and validation study. *JMIR Med Inform*. 2022;10(2):e34492.

32. Dopelt K, Houminer-Klepar N. The impact of social media on disordered eating: Insights from Israel. *Nutrients*. 2025;17(1):180.

33. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. *BMC Med Educ*. 2022;22(1):382.

34. Dünki A, Öztürk Ö. Assessment of informational quality and credibility of YouTube content on osteosarcoma. *J Baltalimanı*. 2025;1(2):53-57.

35. Lin J, Portingale J, Krug I. Virtual feasts and their impact: A pilot study on the influence of mukbang videos on eating behaviours, body image, and emotional wellbeing. *Body Image*. 2025;56:102004.

36. Ataguba G, Kalu I, Chan G, Orji R. Food mukbang on social media: Towards an AI-driven persuasive interventions for living healthy on social media. *AI Soc*. 2025;40:3295-3316.

37. Roorda BA, Molnar M, Cassin SE. “What I eat in a day”: The effect of social media’s diet messaging on body image and dieting intentions. *Dietetics*. 2026;5(1):8.

38. Fiuza A, Rodgers RF. The effects of brief diet and anti-diet social media videos on body image and eating concerns among young women. *Eat Behav*. 2023;51:101811.

39. Serifler S, Gul F. Evaluating tonsillectomy-related YouTube videos via a human expert review and the ChatGPT-4: A multi-method quality analysis. *BMC Med Educ*. 2025;25(1):1157.

Clinical Decision Support via ChatGPT: An Evidence-Based Perspective in Pediatric Occupational Therapy

Gamze Çağla SIRMA*, İbrahim ERARSLAN**, Zeynep BAHADIR***

Abstract

Aim: This study aims to evaluate the accuracy of ChatGPT's responses to clinical questions related to Cerebral Palsy (CP) and Autism Spectrum Disorder (ASD) in the field of occupational therapy, as well as the reliability of the references it provides.

Method: Ten clinical questions were formulated, and answers were prepared based on clinical guidelines and reviewed by two expert clinicians. The same questions were presented to ChatGPT, and its responses and references were rated by two independent evaluators using a four-point Likert scale. Weighted Kappa Coefficients were calculated to assess inter-rater agreement.

Results: ChatGPT's responses demonstrated high accuracy, with strong inter-rater agreement (Weighted Kappa: 0.72–0.75). However, significant deficiencies were identified in reference reliability, with moderate to high inter-rater agreement (Weighted Kappa: 0.75–0.78). Additionally, fictitious reference rates ranged from 40% to 100% for CP-related questions and from 25% to 100% for ASD-related questions.

Conclusion: While ChatGPT shows potential as a clinical decision support tool in occupational therapy, the limitations in reference accuracy restrict its reliability in healthcare applications. Future studies should focus on improving artificial intelligence models' reference accuracy to enhance their applicability in healthcare settings.

Keywords: Artificial intelligence, autism spectrum disorder, cerebral palsy, occupational therapy.

ChatGPT ile Klinik Karar Desteği: Pediatrik Ergoterapide Kanıta Dayalı Bir Perspektif

Öz

Amaç: Bu çalışmanın amacı, Ergoterapi alanında Serebral Palsi (SP) ve Otizm Spektrum Bozukluğu (OSB) ile ilgili klinik sorulara ChatGPT'nin verdiği yanıtların doğruluğunu ve sağladığı kaynakların güvenilirliğini değerlendirmektir.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 02.10.2025 **Kabul / Accepted:** 02.07.2026
DOI: https://doi.org/10.38079/igusabder.1795663
* (Corresponding Author), Lecturer, Department of Occupational Therapy, Faculty of Health Sciences, Fenerbahçe University, Istanbul, Türkiye. E-mail: gamze.dirgen@fbu.edu.tr ORCID https://orcid.org/0000-0003-2976-9035
** Res. Asst., Department of Occupational Therapy, Faculty of Health Sciences, Istanbul Medipol University, Istanbul, Türkiye. E-mail: ibrahim.erarslan@medipol.edu.tr ORCID https://orcid.org/0000-0002-0760-0223
*** Assoc. Prof., Department of Occupational Therapy, Hamidiye Faculty of Health Sciences, University of Health Sciences, Istanbul, Türkiye. E-mail: zeynep.bahadir@sbu.edu.tr ORCID https://orcid.org/0000-0002-7674-9830

ETHICAL STATEMENT: This study was previously presented as an oral presentation at the International Congress on Digitalization and Artificial Intelligence in Healthcare, held at Sinop University, Türkiye, on 21–23 September 2025, and was published in the abstract book (p. 59; e-ISBN: 978-625-96460-0-8).

Yöntem: On klinik soru oluşturulmuş, yanıtlar klinik rehberlere dayalı olarak hazırlanmış ve iki uzman klinisyen tarafından gözden geçirilmiştir. Aynı sorular ChatGPT'ye yöneltilmiş, yanıtları ve kaynakları iki bağımsız değerlendirici tarafından dört dereceli Likert ölçeği ile puanlanmıştır. Değerlendiriciler arası uyumu belirlemek için Ağırlıklı Kappa Katsayıları hesaplanmıştır.

Bulgular: ChatGPT'nin yanıtları yüksek doğruluk göstermiş ve değerlendiriciler arası güçlü uyum saptanmıştır (Ağırlıklı Kappa: 0,72–0,75). Ancak, kaynak güvenilirliğinde önemli eksiklikler bulunmuş olup, değerlendiriciler arası uyum orta-yüksek düzeydeydi (Ağırlıklı Kappa: 0,75–0,78). Ayrıca, hayali kaynak oranları SP ile ilgili sorularda %40–100, OSB ile ilgili sorularda %25–100 arasında değişmiştir.

Sonuç: ChatGPT, ergoterapi alanında klinik karar destek aracı olarak potansiyel göstermektedir; ancak kaynak doğruluğundaki sınırlılıklar, sağlık hizmetlerindeki güvenilirliğini kısıtlamaktadır. Gelecek çalışmalarda, yapay zekâ modellerinin kaynak doğruluğunun geliştirilmesine odaklanılması, sağlık alanındaki uygulanabilirliklerini artıracaktır.

Anahtar Sözcükler: Yapay zekâ, otizm spektrum bozukluğu, serebral palsi, ergoterapi.

Introduction

Artificial intelligence (AI) refers to software designed to perform complex tasks that typically require human intelligence, such as problem-solving, learning, and decision-making[1]. Recent advances in machine learning architectures and the availability of large training datasets have enabled AI systems to perform at levels comparable to humans in specific domains. Among these, ChatGPT, developed by OpenAI, has emerged as a prominent example of AI technology. This AI-powered chatbot utilizes a Generative Pre-training Transformer (GPT) model, which enables it to generate human-like, context-sensitive text based on vast amounts of publicly available data[2]. ChatGPT's ability to perform diverse tasks—ranging from answering questions to summarizing content and generating creative compositions—has contributed to its widespread adoption, with over 100 million users in just two months of its release[3].

In the healthcare sector, ChatGPT has demonstrated the potential to support clinical workflows by synthesizing vast amounts of information, offering quick answers to complex questions, and generating personalized recommendations[4-7]. In a fast-paced clinical environment, where time constraints often limit access to current evidence, ChatGPT's ability to relay critical, evidence-based information could be transformative. By integrating and contextualizing data from sources such as case reports, scientific articles, and clinical guidelines, ChatGPT may enable clinicians to develop individualized treatment plans more efficiently[4,8,9]. However, the utility of ChatGPT in healthcare is not without limitations. Particularly in areas with limited published evidence or where clinical scenarios are highly complex, ChatGPT's responses may lack the necessary depth or accuracy. Furthermore, the phenomenon of "hallucinations," where the model generates fabricated or misleading information, poses significant risks in clinical decision-making[4,10].

One of the core principles of modern healthcare is evidence-based practice (EBP), which involves integrating the best available evidence, clinical expertise, and patient values into the decision-making process. In healthcare settings, EBP is essential for optimizing patient outcomes and ensuring the safe, effective delivery of care[11]. While ChatGPT's design as a "task-agnostic" tool allows it to operate across various disciplines, its capacity to deliver evidence-based insights tailored to specific medical domains remains an area of active investigation[4]. Evaluating the reliability and validity of its responses is particularly crucial for its adoption in the rehabilitation field, where personalized care plans depend heavily on current and accurate evidence.

Occupational therapy (OT) is a rehabilitation field that prioritizes holistic, patient-centered care. Integrating AI tools like ChatGPT into pediatric occupational therapy, the most prevalent area of practice, presents significant opportunities. Recent research has explored various applications of AI in this domain.

For instance, Sharma (2024) highlighted AI's potential to enhance the assessment and intervention processes for children with developmental needs[12]. Similarly, Berrezueta-Guzman et al. (2024) emphasized the effectiveness of AI-based systems in improving evaluation and therapeutic outcomes in pediatric OT[13]. Despite its potential, there is a lack of research evaluating the use of ChatGPT in pediatric OT, particularly in terms of its alignment with evidence-based practice principles. This study aimed to evaluate the evidence-based accuracy of ChatGPT in pediatric occupational therapy for Autism Spectrum Disorder (ASD) and Cerebral Palsy (CP).

Material and Methods

Question Development and Expert Review

Researchers designed ten clinical questions (five for CP and five for ASD) based on relevant clinical guidelines and scientific literature[14-16]. These questions aimed to assess both theoretical knowledge and practical applications. To ensure a balanced evaluation, three questions for each condition focused on general clinical knowledge, while the remaining two were designed as case-based scenarios simulating real-world clinical decision-making (see supplementary). The entire process, including question formulation, expert review, ChatGPT evaluation, and data analysis, was conducted in English.

The initial answer key for the formulated questions was developed based on clinical guidelines and expert recommendations. Two independent researchers reviewed the prepared responses, providing feedback for necessary modifications and refinements. After incorporating their recommendations, the revised answer key was further evaluated and finalized by two experts, each with at least five years of clinical experience (≥5 years of practice).

Data Collection Process

A detailed planning phase was conducted to determine the optimal way to structure the questions for ChatGPT. During this phase, prompt formats were developed to ensure accurate and comprehensive responses. To optimize the phrasing of the queries, ChatGPT was consulted by asking, "*What should we pay attention to in order to get the right answers when asking you questions for professional information in the field of occupational therapy?"* The following responses, along with explanations, were received, and the questions were structured accordingly: be specific in your inquiry, clarify the context, ask for evidence-based answers, specify frameworks or models, include cultural or regional context, state your role and purpose, ask follow-up questions, request examples or practical applications, be clear about the level of detail required, and cross-verify information.

Based on these principles, the final question set was structured and optimized. To ensure transparency and reproducibility, an example of the prompt structure used in the study is presented below:

"As a pediatric occupational therapist, I will ask some questions. 1. Which interventions can I use to improve motor development in children with CP diagnosis or CP risk in the 0-5 age range? In the text, add the scientific reference for this information."

The questions were then entered into ChatGPT (GPT-4o, OpenAI) in January 2025, and the generated responses with references were systematically recorded. ChatGPT provided the sources in APA format, including years of publication.

Evaluation of Responses and References

The accuracy of the responses and the validity of the references were independently evaluated by two raters with at least five years of clinical experience. Both raters were clinician-academics and were familiar with evidence-based practice, literature searching, and verification of scientific sources using major academic databases such as PubMed and Scopus. Each response was rated using a four-point scale (4= Completely correct, 3= Almost correct, 2= Partially correct, 1= Completely incorrect).

The validity of the references was assessed based on the APA-formatted citations and publication years provided by ChatGPT. If a cited article or source did not match the given bibliographic details, additional searches were conducted. References that could not be located through any means were classified as "fictitious references".

Ethical Statement

Ethical approval was not required for this study because it did not involve human participants, patient records, biological samples, or animal subjects. The study evaluated responses generated by ChatGPT (GPT-4o, OpenAI) to researcher-developed clinical questions and assessed their accuracy and reference validity against evidence-based clinical guidelines. No personal, confidential, or identifiable data were collected or analyzed. Consequently, informed consent was not applicable.

Data Analysis

The number of 'fictitious references' was calculated as a percentage. Weighted Kappa Coefficients were also calculated to assess inter-rater reliability. Results were expressed as median (Md) and interquartile range (IQR). The Weighted Kappa Coefficient is a statistical method used to measure the agreement between two or more raters. This method assigns different weights according to the severity of errors and is particularly effective for ordinal data, as it captures the significance of disagreements among raters.

The interpretation of Weighted Kappa Coefficient values followed standard classification. A coefficient below 0.40 indicated poor agreement, values between 0.41 and 0.60 represented moderate agreement, values ranging from 0.61 to 0.80 were classified as good agreement, and values between 0.81 and 1.00 denoted excellent agreement[17].

Results

The evaluation of answer accuracy and reference accuracy was conducted separately for CP- and ASD-related questions. The findings are summarized in Table 1 and Table 2.

Table 1. Summary of scores for output content, reference accuracy, and weighted kappa coefficient

Category	**CP**		**ASD**	
	Assessor 1	**Assessor 2**	**Assessor 1**	**Assessor 2**
Accuracy of the generated output	4.0 [3.0–4.0]	4.0 [3.0–4.0]	4.0 [3.0–4.0]	4.0 [3.0–4.0]
Weighted kappa coefficient of the generated output	**0.72***	**0.72***	**0.75***	**0.75***
Accuracy of the generated references	2.0 [1.0–3.0]	2.0 [1.0–3.0]	2.0 [1.0–3.0]	2.0 [1.0–3.0]
Weighted kappa coefficient of the generated references	**0.75***	**0.75***	**0.78***	**0.78***

*Note: Statistically significant ($p< 0.05$).

Table 2. Reference accuracy of different occupational therapy questions

Question No	Question Description			Fictitious / Total References	Rate of Fictitious References
	Type	Condition	Focus		
1.	General	CP	OT Role	7/7	%100
2.	General	CP	OT Interventions	7/7	%100
3.	General	CP	OT Approaches	5/5	%100
4.	Clinical Case	CP	Assessment	6/15	%40
5.	Clinical Case	CP	Intervention	7/7	%100
6.	General	ASD	OT Role	7/8	%87
7.	General	ASD	Interventions	6/9	%66
8.	General	ASD	Approaches	7/11	%81
9.	Clinical Case	ASD	Assessment	2/12	%25
10.	Clinical Case	ASD	Intervention	7/12	%58

Accuracy of the Generated Output

For CP-related questions (Questions 1–5), the median accuracy score was 4.0 [3.0–4.0], with an IQR of 1.0. The weighted kappa coefficient was 0.72, indicating moderate to high agreement. Questions 1 and 5 showed full agreement (kappa= 1.00), whereas Questions 2, 3, and 4 exhibited some variability.

For ASD-related questions (Questions 6–10), the median accuracy score remained 4.0 [3.0–4.0], with an IQR of 1.0. The weighted kappa coefficient was 0.75, reflecting a high level of agreement. Full agreement was observed for Questions 6, 8, and 9, while Questions 7 and 10 showed some discrepancies between assessors.

Accuracy of the Generated References

For CP-related references, the median accuracy score was 2.0 [1.0–3.0], with an IQR of 1.0. The weighted kappa coefficient was 0.75, indicating high inter-rater agreement.

However, the relatively low median score suggests that references were frequently rated as partially correct (2) or completely incorrect (1), raising concerns about their reliability.

For ASD-related references, the median accuracy score was also 2.0 [1.0–3.0], with an IQR of 1.0. The weighted kappa coefficient was 0.78, showing a strong level of agreement. While Questions 6, 7, 8, and 9 demonstrated consistent assessments, Question 10 showed some variability between assessors.

Table 2 provides additional insight into reference accuracy by detailing the proportion of fictitious references per question. Among CP-related questions, fictitious reference rates ranged from 40% (Question 4) to 100% (Questions 1, 2, and 5). In ASD-related questions, the fictitious reference rate varied from 25% (Question 9) to 100% (Question 3). In particular, general knowledge questions (e.g., Questions 1, 2, 3, 6, 7, and 8) had a higher percentage of fictitious references compared to case-based assessment questions (e.g., Questions 4, and 9).

Overall Assessment

The results indicate that the generated output was generally rated as "completely correct" (4) or "almost correct" (3), confirming its overall accuracy. The weighted kappa values (0.72–0.75) suggest that assessors maintained moderate to high consistency in evaluating the responses.

In contrast, reference accuracy was lower, with most references classified as "partially correct" (2) or "completely incorrect" (1). Despite the high inter-rater agreement (weighted kappa= 0.75– 0.78), the frequent inaccuracies in references—particularly in CP-related responses—highlight the need for stricter validation criteria.

Discussion

This study aimed to assess the accuracy of ChatGPT’s responses to clinical questions related to CP and ASD in the field of occupational therapy, as well as the reliability of the references it provided. The results indicate that while ChatGPT's responses were generally accurate and aligned with clinical guidelines, the reliability of the references was significantly lower. The inter-rater agreement was moderate to high, with weighted kappa coefficients ranging from 0.72 to 0.75 for response accuracy and 0.75 to 0.78 for reference accuracy. These findings suggest that although ChatGPT can serve as a valuable tool for retrieving clinical information, its ability to generate reliable references remains a critical limitation.

The accuracy of ChatGPT’s responses was found to be high, with assessors generally rating them as clinically sound and relevant to the questions posed. These results align with previous studies examining AI-powered decision support in healthcare. Scaff et al. (2025) assessed the performance of large language models in answering patient inquiries about low back pain and found that, while AI-generated responses were informative, inconsistencies in contextual depth and phrasing were noted[18]. Similarly, Sawamura et al. (2024) analyzed ChatGPT's responses to physiotherapy-related clinical questions and reported that, while the information was mostly accurate, minor discrepancies existed between AI-generated content and established best practices[19]. These findings suggest that ChatGPT can synthesize and relay relevant clinical information efficiently, though limitations remain in the precision and specificity of certain answers.

The use of AI-based clinical decision support systems is gaining traction due to their ability to rapidly process vast amounts of medical information[20]. However, as emphasized by Akalin and Veranyurt (2022), reliance on AI-generated responses without expert oversight poses risks, particularly in nuanced clinical decision-making[21]. In this study, high inter-rater agreement was observed in evaluating response accuracy, suggesting that the generated content was consistently aligned with clinical guidelines. Nonetheless, discrepancies in the assessment of certain questions highlight the need for further validation, particularly in complex cases requiring individualized decision-making.

Despite the high accuracy of responses, a major limitation identified in this study was the reliability of ChatGPT's references. Such inaccuracies are commonly referred to as "hallucinations" in large language models. LLMs such as ChatGPT function primarily as probabilistic next-token prediction systems, generating text based on patterns learned from large training corpora rather than retrieving information directly from indexed academic databases. Consequently, when asked to provide references, the model may generate citations that appear plausible in structure but do not correspond to real publications, particularly when real-time access to bibliographic databases is not available. Consistent with this mechanism, a significant proportion of the references in the present study were found to be incomplete, inaccurate, or entirely fabricated, limiting their applicability in evidence-based practice. These findings are consistent with prior studies investigating AI-generated references in scientific and medical contexts. Graf et al. (2024) assessed the reference accuracy of ChatGPT, Google Bard, and Chatsonic in generating citations for randomized controlled trials in obstetric literature, reporting that Google Bard achieved the highest reference accuracy at only 13.6%, while ChatGPT and Chatsonic exhibited significantly lower accuracy rates of 2.4% and 0%, respectively[22]. Similarly, Sawamura et al. (2024) noted that ChatGPT frequently provided nonexistent or misinterpreted references in physiotherapy-related clinical questions, reinforcing concerns about AI-generated citations[19].

The analysis of fictitious reference rates indicates a pattern based on question type. The lowest rates of fictitious references were observed in case-based clinical questions (e.g., Question 9: 25%), while the highest rates were found in general knowledge questions (e.g., Questions 1, 2, and 3: 100%). This suggests that ChatGPT's reference reliability is more consistent when responding to case-based scenarios requiring structured clinical reasoning, whereas general queries elicit a higher proportion of inaccurate citations. This discrepancy may be attributed to the model's reliance on broad textual patterns rather than direct retrieval of peer-reviewed sources. Additionally, these findings suggest that clinicians may benefit from formulating more detailed and specific questions, as structured queries appear to yield more reliable references and clinically relevant responses.

The limitations of AI in generating reliable references may stem from its reliance on static training datasets that do not allow real-time access to up-to-date, peer-reviewed literature[23]. This issue is further compounded by the tendency of large language models to generate citations based on linguistic patterns rather than direct retrieval from indexed medical databases[24]. Kunt (2023) also highlights that the effectiveness of AI-driven tools in healthcare is directly linked to the accuracy and reliability of the information they provide[25]. Given these limitations, ensuring proper validation of AI-generated content remains a crucial requirement for safe clinical integration.

Several limitations should be acknowledged. This study evaluated only a specific version of ChatGPT, without comparisons to alternative AI models such as Google Bard or Claude. Given the rapid development of AI technologies, newer versions may demonstrate improved reference accuracy and response precision. Furthermore, the study was limited to CP- and ASD-related clinical questions, which may not fully capture the model's performance across broader occupational therapy domains. Additionally, the relatively small number of questions included in the study (n=10) may limit the statistical power and generalizability of the findings. The assessment was conducted by expert raters, which enhances reliability but may introduce subjective biases. Additionally, the study did not analyze ChatGPT's ability to adapt to different question styles or prompts, which could influence response accuracy. Furthermore, the questions and evaluations were conducted in English, which may have contributed positively to the accuracy of ChatGPT's responses but also presents a limitation, as the raters were not native English speakers. In addition, it is not clear how asking in the native language would have affected the accuracy of the ChatGPT responses.

Several recommendations could be made for future research. First, conducting comparative analyses between different versions of ChatGPT and other AI-based models would provide a more comprehensive understanding of their strengths and weaknesses. Additionally, investigating how ChatGPT responds to different question styles across various diagnostic groups could offer valuable insights into its ability to generate clinically relevant information tailored to specific conditions. Furthermore, developing a fully evidence-based ChatGPT-Academic model designed to support clinical decision-making could enhance its applicability in healthcare settings. Finally, enhancing AI models through training with reliable literature sources and developing more robust verification mechanisms would be essential in improving the accuracy and trustworthiness of AI-generated content in healthcare applications.

Disclosure statement: No potential conflict of interest was reported by the author(s).

Funding: This research received no external funding.

REFERENCES

1. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. *arXiv*. 2023. doi: 10.48550/arXiv.2303.13375.

2. Liu Y, Han T, Ma S, et al. Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*. 2023;100:017.

3. Hu K. ChatGPT sets record for fastest-growing user base - analyst note. *Reuters*. February 2, 2023. Accessed February 1, 2023. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01.

4. Campbell WA IV, Chick JF, Shin D, Makary MS. Understanding ChatGPT for evidence-based utilization in interventional radiology. *Clin Imaging*. 2024;110:098.

5. Rao A, Pang M, Kim J. Assessing the utility of ChatGPT throughout the entire clinical workflow. *medRxiv*. 2023.

6. Garg RK, Urs VL, Agarwal AA, Chaudhary SK, Paliwal V, Kar SK. Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review. *Health Promot Perspect*. 2023;13(3):183.

7. Von Ende E, Ryan S, Crain MA, Makary MS. Artificial intelligence, augmented reality, and virtual reality advances and applications in interventional radiology. *Diagnostics*. 2023;13(5):892.

8. Greengard S. ChatGPT: Understanding the ChatGPT AI chatbot. *Eweek*. December 9, 2022. https://www.eweek.com/big-data-and-analytics/chatgpt/.

9. Manning CD. Human language understanding & reasoning. *Proc Am Acad Arts Sci*. 2022;151(2):127-138. doi: 10.1162/daed_a_01905.

10. Mündler N, He J, Jenko S, Vechev M. Self-contradictory hallucinations of large language models: Evaluation, detection, and mitigation. *arXiv*. 2023.

11. Hoffmann T, Bennett S, Del Mar C. *Evidence-based practice across the health professions*. Philadelphia, PA: Elsevier Health Sciences; 2023.

12. Sharma N. Use of AI in pediatric occupational therapy: A review. *Rehabil Sci*. 2024;9(2):21-26. doi: 10.11648/j.rs.20240902.12

13. Berrezueta-Guzman S, Kandil M, Martín-Ruiz ML, de la Cruz IP, Krusche S. Exploring the efficacy of robotic assistants with ChatGPT and Claude in enhancing ADHD therapy: Innovating treatment paradigms. In: *2024*

International Conference on Intelligent Environments (IE). 2024:25-32. doi: 10.1109/IE61493.2024.10599903.

14. Morgan C, Fetters L, Adde L, et al. Early intervention for children aged 0 to 2 years with or at high risk of cerebral palsy: International clinical practice guideline based on systematic reviews. *JAMA Pediatr*. 2021;175(8):846-858. doi: 10.1001/jamapediatrics.2021.0878.

15. Steinbrenner JR, Hume K, Odom SL, et al. *Evidence-based practices for children, youth, and young adults with autism*. Chapel Hill, NC: The University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Institute, National Clearinghouse on Autism Evidence and Practice Review Team; 2020.

16. Cahill SM, Beisbier S. Occupational therapy practice guidelines for children and youth ages 5–21 years. *Am J Occup Ther*. 2020;74(4):7404397010p1-7404397010p48.

17. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology*. 2003;228:303-308.

18. Scaff SP, Reis FJ, Ferreira GE, Jacob MF, Saragiotto BT. Assessing the performance of AI chatbots in answering patients' common questions about low back pain. *Ann Rheum Dis*. 2025;84(1):143-149.

19. Sawamura T, Lee H, Martinez R. Assessing the reliability of ChatGPT-generated references in physiotherapy clinical decision-making. *Physiother Res J*. 2024;32(2):89-101.

20. Kaplan M, Çakar F, Bingöl H. Sağlık alanında yapay zekânın kullanımı: Derleme. *Muş Alparslan Univ Sağlık Bilim Derg*. 2024;4(3):75-85.

21. Akalin B, Veranyurt Ü. Sağlık bilimlerinde yapay zekâ tabanlı klinik karar destek sistemleri. *Gevher Nesibe J Med Health Sci*. 2022;7(18):64-73.

22. Graf A, Smith B, Johnson C. Evaluating the accuracy of AI-generated references in obstetric randomized controlled trials: A comparative study of ChatGPT, Google Bard, and Chatsonic. *J Obstet Res*. 2024;45(3):214-226.

23. Kim S. In the era of ChatGPT, can medical artificial intelligence replace the doctor? *Korean J Med*. 2023;98(3).

24. Waisberg E, Ong J, Masalkhi M, et al. GPT-4: A new era of artificial intelligence in medicine. *Ir J Med Sci*. 2023.

25. Kunt MM. Tıpta dil tabanlı yapay zeka kullanımı. *Anatolian J Emerg Med*. 2023;6(3):137-140.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 65-79.

Dietitians' Views on Artificial Intelligence: An In-Depth Study*

Elif GÜNER**, Kübra YUCA***, Emel ÖZTÜRK****, Berfin BALCIOĞLU*****, Zeynep İrem KÜÇÜKÇAKAN******

Abstract

Aim: Today, the use of artificial intelligence technologies is becoming increasingly common in many areas from health services to nutrition planning. The aim of this study is to explore dietitians' perceptions, experiences, and professional perspectives on the role of artificial intelligence in the field of nutrition.

Method: In this qualitative study, online in-depth interviews were conducted with 29 dietitians working in different sectors between March 27 and April 28, 2025. In the interviews lasting 15-20 minutes, a semi-structured topic guide consisting of 10 questions was used. Transcripts were analyzed using the thematic analysis method.

Results: Dietitians' statements were evaluated under the themes of (1) professional transformation and the role of artificial intelligence, (2) ethics, trust and professional responsibility, (3) professional competence, education and adaptation, (4) Scientific validity, reliability and personalization. Dietitians evaluated artificial intelligence as a helpful tool that saves time and facilitates routine tasks, but emphasized that human interaction, empathetic communication, and ethical responsibility cannot be replaced by technology. Although it is thought that artificial intelligence will lead to a radical change in the field of nutrition, there are concerns about issues such as patient privacy, ethical values, and lack of scientific evidence.

Conclusion: To our knowledge, this is the first study in Türkiye to explore dietitians' perspectives on artificial intelligence in nutrition. This study shows that artificial intelligence presents both risks and opportunities in the dietitian profession, and draws attention to the fact that professionals should be equipped in terms of knowledge, ethics and digital competence in this transformation process.

Keywords: Dietetics, artificial intelligence, qualitative research, ethics, digital health.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 17.03.2026 **Kabul / Accepted:** 16.06.2026
DOI: https://doi.org/10.38079/igusabder.1782791
* The present study was derived from a project funded by the TÜBİTAK 2209-A Undergraduate Research Support Program, entitled "Qualitative Analysis of Dietitians' Views on the Role of Artificial Intelligence in Nutrition" (1919B012467807).
** Asst. Prof., Istanbul Sabahattin Zaim University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: dyt.elifemiroglu@gmail.com ORCID https://orcid.org/0000-0003-0575-5587
*** Graduated Dietician, Istinye University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: yuca14kbr@hotmail.com ORCID https://orcid.org/0009-0006-6533-5815
**** Graduated Dietician, Istinye University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: emelcozovic@gmail.com ORCID https://orcid.org/0009-0008-0027-7700
***** Graduated Dietician, Istinye University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: berfinbalciogluu@icloud.com ORCID https://orcid.org/0009-0002-3998-6367
****** Graduated Dietician, Istinye University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: iremkc@hotmail.com ORCID https://orcid.org/0009-0007-8022-6892

ETHICAL STATEMENT: This study was carried out with the approval of Istinye University Ethics Committee for Research on Social and Human Sciences, date 30/12/2024 and numbered 2024-12. A signed subject consent form in accordance with the Declaration of Helsinki was obtained from each participant.

Diyetisyenlerin Yapay Zekâ Hakkındaki Görüşleri: Derinlemesine Bir Çalışma

Öz

Amaç: Yapay zekâ teknolojilerinin kullanımı günümüzde sağlık hizmetlerinden beslenme planlamasına kadar birçok alanda giderek yaygınlaşmaktadır. Bu çalışmanın amacı, diyetisyenlerin yapay zekânın beslenme alanındaki rolüne ilişkin algılarını, deneyimlerini ve mesleki bakış açılarını araştırmaktır.

Yöntem: Bu nitel çalışmada, farklı sektörlerde çalışan 29 diyetisyenle 27 Mart - 28 Nisan 2025 tarihleri arasında çevrimiçi derinlemesine görüşmeler gerçekleştirilmiştir. 15-20 dakika süren görüşmelerde, 10 sorudan oluşan yarı yapılandırılmış bir konu rehberi kullanılmıştır. Transkriptler tematik analiz yöntemi ile analiz edilmiştir.

Bulgular: Diyetisyenlerin ifadeleri (1) mesleki dönüşüm ve yapay zekânın rolü, (2) etik, güven ve mesleki sorumluluk, (3) mesleki yeterlilik, eğitim ve adaptasyon, (4) bilimsel geçerlilik, güvenilirlik ve kişiselleştirme temaları altında değerlendirilmiştir. Diyetisyenler, yapay zekâyı zamandan tasarruf sağlayan ve standart görevleri kolaylaştıran faydalı bir araç olarak değerlendirmiş olsalar da, insan dokunuşunun, empatik iletişimin ve etik sorumlulukların teknolojiyle değiştirilemeyeceğini vurgulamaktadırlar. Yapay zekânın beslenme alanında köklü bir değişime yol açacağı düşünülse de, hasta mahremiyeti, etik değerler ve bilimsel kanıt eksikliği gibi endişeler bulunmaktadır.

Sonuç: Bildiğimiz kadarıyla, bu çalışma Türkiye'de diyetisyenlerin beslenme alanında yapay zekâya bakış açılarını inceleyen ilk çalışmadır. Çalışmamız, yapay zekânın diyetisyenlik mesleğinde bazı risklerin yanı sıra fırsatlar da sunduğunu göstermekte ve bu dönüşüm sürecinde profesyonellerin bilgi, etik ve dijital yetkinlik açısından donanımlı olması gerektiğine dikkat çekmektedir.

Anahtar Sözcükler: Diyetetik, yapay zekâ, niteleyici araştırma, etik, dijital sağlık.

Introduction

Artificial intelligence (AI) was first mentioned as a term in 1956, covering a variety of applications, from general applications such as learning to specific areas such as playing chess, writing poetry, or diagnosing a disease. AI was defined by American computer scientist McCarthy in 2007 as the engineering and science of making intelligent machines, including intelligent computer programs[1]. In addition to this wide range of uses, the use of AI in the field of health is becoming increasingly common. There is a growing number of studies evaluating the use of AI in different fields such as assessment of fracture risk[2], psychiatry[3], early diagnosis of neonatal sepsis[4], perioperative medicine[5], new drug discovery[6], medical sensor technology[7], nursing practices[8], obstetric ultrasonography[9].

In the field of nutrition and dietetics, there are studies examining the role of AI in a wide range of areas, such as the estimation of the nutrient content of meals, pediatric nutrition, weight control, oncology nutrition, nutrition in diabetes, hypertension, and metabolic syndrome[10-18]. Dietitians can use AI systems to help society eat healthily, create more effective nutrition education programs, and disseminate accurate nutritional information[19]. However, it is reported that AI is less researched and used in the field of nutrition and dietetics compared to fields such as medicine, that the information provided by AI has not been sufficiently validated, that there is a need for clinical trials examining the effectiveness of interventions using AI, and that ethical violations are a major concern[20].

Despite the growing global interest in the use of AI in healthcare, there is a lack of research examining how dietitians in Türkiye perceive and integrate these technologies into their professional practice. Understanding the perspectives of dietitians in this context is important, as cultural, educational, and health system-related factors may shape their views differently compared to other countries. This qualitative study aimed to explore the views and experiences of dietitians on the use of AI in nutrition and dietetics.

Material and Methods

Study Design

This qualitative study was designed to explore dietitians' views on the role of AI in nutrition. In this context, the study aimed to better understand and make sense of dietitians' perceptions, experiences and perspectives on AI.

Sample

Participants were selected from among the dietitians who graduated from the Departments of Nutrition and Dietetics in Türkiye and are actively working. Participants who volunteered for the study were selected by the maximum diversity sampling method. Participants were selected from different professional experience and age groups. Thus, the study aimed to obtain comprehensive and rich data with different experiences and perspectives.

In qualitative research, data saturation is defined as the inability to add a new theme, concept or finding to the analysis process when new data is collected. In this study, when the number of participants reached twenty, repetitions in themes, sub-themes and codes began to be seen intensively, and the data obtained from the new participant did not contribute to the existing themes. Thus, data saturation was considered to have been reached. At the end of the coding process, it was determined that the existing themes were sufficient and rich enough to meet the research questions.

In-Depth Interviews

Interviews were conducted via online platforms between March 27 and April 28, 2025, on the pre-planned day and time according to the availability of the participants. During the interviews, audio recordings were taken with the permission of the participants. The interviews lasted between 15 and 20 minutes.

In the study, a data collection form consisting of 6 questions and a semi-structured interview guide consisting of 10 questions were used as data collection tools. The data collection form includes questions about the dietitian's age, gender, sector, experience level, educational level and graduation year. The semi-structured interview guide includes the following questions:

1. *Are you taking advantage of AI? If so, can you explain in which areas and for what purpose you use it?*

2. *What do you think are the advantages and disadvantages of using AI?*

3. *Do you think AI can replace dietitians, or do you think that the observation and experience of dietitians will always be needed?*

4. *What are your thoughts on the claim that AI can provide more accurate and personalized recommendations in the field of nutrition and dietetics?*

5. *How can dietitians improve themselves to work with AI?*

6. *What do you think about the ethical use of AI?*

7. *How will the integration of AI into nutrition and dietetics practices impact your daily workflow and patient management processes?*

8. *What do you think about the effects of the possibility of carrying out nutrition counseling processes with AI-based systems on dietitians in the future?*

9. *How can the spread of AI-based nutrition guides and mobile applications affect dietitians?*

10. *Is there anything else you want to say?*

Ethical Statement

Participant responses were anonymized in a way that did not contain any personal data, taking into account ethical principles, and codes (P1, P2, etc.) were assigned to each participant. All participants were informed and their consent was obtained before the interviews. The study was carried out in accordance with the Declaration of Helsinki. Ethics committee permission was obtained from Istinye University Social Sciences and Humanities Ethics Committee (Meeting No: 2024-12, Meeting Date: 30.12.2024).

Data Analysis

Descriptive statistics (mean, standard deviation, minimum, maximum) were calculated for participant characteristics.

Thematic analysis approach was used in the analysis of qualitative data. Thematic analysis is a scientifically accepted qualitative analysis technique that allows certain patterns, meanings and themes to be systematically revealed from answers to openended questions. The thematic analysis was conducted following the phases proposed by Braun and Clarke. Accordingly, the analysis involved (I) repeated reading of the transcripts to achieve data familiarization, (II) open coding of meaningful units, (III) grouping similar codes into categories, and (IV) organizing these categories into overarching themes. An inductive, data-driven approach was adopted in the development of themes, and the consistency between codes and themes was checked iteratively throughout the analysis process[21]. MAXQDA software (VERBI Software, Berlin, Germany) was used to facilitate the systematic coding of transcripts, organization of codes, and development of themes during the thematic analysis process. The coded data were organized into main themes and sub-themes, taking into account the similarity of content and repeated patterns. An inductive approach was adopted in the creation of themes. In order to ensure reliability in the coding and theming process, the obtained codes and themes were constantly checked. In the Results section, citations are included for each main theme and sub-theme. Thus, the verifiability of the findings was increased. Care has been taken to ensure researcher impartiality in order to prevent possible biases during the analysis process. The coding process was carried out by one researcher. This decision was based on the fact that the same researcher was involved in the development of the interview guide, data collection, and analysis, which helped preserve contextual coherence and interpretive consistency.

Results

In this study, which aimed to explore dietitians' perceptions, experiences and professional perspectives on the role of AI in the field of nutrition, in-depth interviews were conducted with a total of 29 (female= 24, male= 5) dietitians. 37.93% of the participants have a bachelor's degree, 58.62% have a master's degree and 3.45% have a doctoral degree. The ages of dietitians ranged from 23 to 48 years, with a median age of 28.14±6.08 years. The graduation years of the participants are distributed between 2012 and 2024. The years of professional experience of dietitians vary between 1 and 13 years, and the average years of professional experience is 4.14±3.40 (Table 1).

Table 1. Characteristics of participants

Participant	Age	Gender	Educational Level	Graduation Year	Years of Professional Experience	Institution
P1	24	Female	Master's Degree	2023	1	Online Counseling
P2	32	Male	Master's Degree	2014	10	Counseling Center
P3	25	Female	Bachelor's Degree	2023	1	Catering
P4	28	Female	Master's Degree	2020	4	University

P5	23	Female	Bachelor's Degree	2024	1	Catering
P6	23	Female	Master's Degree	2023	2	Counseling Center
P7	26	Female	Bachelor's Degree	2021	4	Private Hospital
P8	24	Male	Bachelor's Degree	2024	1	Online Counseling
P9	25	Female	Bachelor's Degree	2021	4	Catering
P10	24	Female	Bachelor's Degree	2024	1	Private Hospital
P11	26	Female	Master's Degree	2024	4	Private Hospital
P12	27	Female	Master's Degree	2020	4	University
P13	38	Male	Master's Degree	2015	9	State Hospital
P14	25	Female	Master's Degree	2022	3	Counseling Center
P15	33	Female	Master's Degree	2016	9	Private Hospital
P16	23	Female	Bachelor's Degree	2024	1	Online Counseling
P17	25	Female	Master's Degree	2021	3	Catering
P18	37	Female	Doctoral Degree	2012	13	Counseling Center
P19	27	Female	Master's Degree	2020	5	Private Hospital
P20	24	Female	Bachelor's Degree	2022	3	Counseling Center
P21	26	Female	Bachelor's Degree	2022	2	Private Hospital
P22	26	Female	Master's Degree	2022	3	Private Hospital
P23	25	Female	Master's Degree	2022	3	Private Hospital
P24	48	Male	Master's Degree	2021	4	Private Hospital

P25	35	Female	Master's Degree	2023	2	Counseling Center
P26	36	Female	Master's Degree	2014	10	Municipality
P27	23	Female	Master's Degree	2023	2	Counseling Center
P28	23	Female	Bachelor's Degree	2024	1	Counseling Center
P29	35	Male	Bachelor's Degree	2015	10	Counseling Center

P: Participant.

The answers given by the participants to the semi-structured questions were evaluated by thematic analysis method and four main themes and eight sub-themes in Table 2 were created.

Table 2. Themes and sub-themes

Themes	**Sub-themes**
1. Professional Transformation and the Role of AI	1.1. The Future of the Dietitian Profession and Role Change
	1.2. Automation of Routine Work and Time Management
2. Ethics, Trust and Professional Responsibility	2.1. Data Security and Patient Privacy
	2.2. Limits of Liability in Decision Support Systems
3. Professional Competence, Training and Adaptation	3.1. New Competencies and the Need for Continuous Training
	3.2. Professional Adjustment, Resistance, and Fears
4. Scientific validity, Reliability and Personalization	4.1. Scientific Basis and Application Reliability
	4.2. Personalized Counseling

AI: Artificial Intelligence

Theme 1. Professional Transformation and the Role of AI

Sub-theme 1.1. The Future of the Dietitian Profession and Role Change

The vast majority of dietitians agree that AI-based systems will radically transform the dietitian profession. Some dietitians have emphasized that while AI can provide significant convenience in routine tasks, it is not possible to completely eliminate the human factor.

"I think there will always be a need for the observation and experience of dietitians, because AI cannot account for all variables." (P1)

"I think we can be like extras. However, human contact is always needed. No technology can fully replace humans in decision-making." (P3)

"AI may reduce demand for dietitians, but one-on-one human relationships are still important for some clients." (P4)

Sub-theme 1.2. Automation of Routine Work and Time Management

Saving time and streamlining workflow (as one of AI's most prominent contributions) has often been emphasized.

"Thanks to AI, dietitians can spend more time with their clients and get rid of unnecessary workload." (P7)

"Yes, AI makes diet program creation easy. However, the human factor is still required for detailed evaluation and follow-up." (P5)

Theme 2. Ethics, Trust and Professional Responsibility

Sub-theme 2.1. Data Security and Patient Privacy

A significant portion of dietitians noted that data privacy and patient privacy are critical risks in AI applications. It was emphasized that ethical sensitivities are one of the core values of the profession and that it is imperative to uphold these values in AI applications.

"This is a really important issue. Much more care should be taken, especially regarding the confidentiality of patient information." (P1)

"Data privacy is a must. One of my biggest fears is that AI applications may violate patient privacy." (P18)

"The biggest ethical risk is that data falls into the wrong hands and patients are harmed." (P11)

Sub-theme 2.2. Limits of Liability in Decision Support Systems

Some dietitians have pointed out uncertainties about who will be responsible for the negativities that may arise as a result of the decisions made by AI-powered systems.

"Who will be responsible if there is a mistake? Dietitian, app developer, AI system?" (P15)

"If AI is given carte blanche, we may lose our professional responsibilities. That's why the boundaries need to be drawn clearly." (P20)

Theme 3. Professional Competence, Training and Adaptation

Sub-theme 3.1. New Competencies and the Need for Continuous Training

Dietitians stated that technology literacy and digital competencies should be developed in order to survive professionally in the age of AI. It was also suggested that university curricula should be updated and continuing education programs should be organized.

"Since AI is developing every day, I think a dietitian profile that follows innovations and receives continuous training is essential." (P3)

"It is important for dietitians to receive a basic education on AI, at least to understand how algorithms work." (P9)

"These subjects must be taught as courses in universities." (P17)

Sub-theme 3.2. Professional Adjustment, Resistance, and Fears

Some of the participants stated that they experienced anxiety and resistance to technological transformation. Some participants stated that professional adaptation would only be possible with the right training and guidance.

"Sometimes I'm scared because I feel like my profession is going to go away." (P13)

"You have to be open to every innovation, but it still takes time to get used to it." (P22)

"I strive to improve myself. But some of my friends are very nervous." (P6)

Theme 4. Scientific Validity, Reliability and Personalization

Sub-theme 4.1. Scientific Basis and Application Reliability

A significant portion of dietitians have expressed skepticism about the compatibility of AI-based systems with current scientific knowledge and guidelines. The importance of personalized applications in dietetics was emphasized, and it was stated that standard algorithms may not be suitable for every individual.

"From a scientific point of view, AI is still very new. More evidence and research are needed." (P1)

"The diet should be personalized. AI helps to a certain extent, but it can't detect all the differences." (P21)

Sub-theme 4.2. Personalized Counseling

Some participants highlighted that while AI applications can be particularly beneficial for simple follow-ups, the dietitian's role is indispensable in complex cases.

"For simple follow-ups, yes, but in complex cases, the human eye is essential." (P8)

"AI helps with routine tasks, but professional knowledge is essential for a detailed nutrition plan." (P10)

"Every individual is different; A single suggestion would not be healthy for everyone." (P25)

Discussion

The science of Nutrition and Dietetics has traditionally been based on clinical trials and observational studies. However, AI's capabilities include uncovering complex relationships in large datasets, making it widely applicable to personalized dietary recommendations and disease prediction. Integrating AI-based applications into nutrition and dietetics is thought to reshape nutrition interventions, facilitating technological advancements[22]. In this study, dietitians' views on this growing role of AI in nutrition and dietetics were explored.

Healthcare professionals can benefit from AI to improve healthcare operations, reduce errors, ease the burden on hospitals, and increase quality and sustainability[1]. In nutritional science, AI-based systems can be used as an aid in areas such as screening, disease risk prediction, supporting workflow and management, assessment and intervention[23]. The findings obtained in the in-depth interviews in this study show that dietitians see AI as an aid that provides convenience in routine tasks and reduces unnecessary workload. However, some dietitians have expressed concerns about the growing role of AI in the field of nutrition. Despite its role as an assistant and a facilitator of routine tasks, there are dietitians who believe that AI will bring about a fundamental change in nutritional science. These views are consistent with some other studies that question whether AI can replace different healthcare professionals (dietitians, psychiatrists, etc.) in the future[24,25]. However, it is an undeniable fact that for AI to be successful in clinical use, the expertise of professionals is primarily needed. In particular, the lack of logic in some questions asked or the excess of missing data will make the solutions offered by AI inadequate[26]. The impact of AI on healthcare quality is related to the quality of data entry. Therefore, it is very important for dietitians to participate in the development of AI[27]. For this purpose, integrating the subject of AI into undergraduate education curricula will make a significant contribution to organizing continuing education programs and increasing the knowledge level of dietitians about the limits and functions of AI. In this study, some dietitians emphasized the need for education on AI.

Despite the opportunities, there are also a number of threats to the use of AI systems in the field of nutrition. Failure to consider individual variables, difficulties in dealing with complex scenarios, and lack of transparency of the source of information are some of these threats[23]. AI systems are expected to think and act in a human-like manner, behave rationally, and achieve the best possible outcome. While the integration of AI into healthcare is expected in the future, it must first undergo ethical scrutiny and gain societal acceptance[28]. In this study, some dietitians expressed resistance to the use of AI systems. These insights align with the existing literature, which suggests that the use of AI in nutrition and dietetics is lagging behind compared to fields like medicine. This may be due to dietitians' lack of knowledge on the subject, or it may be a result of their professional and ethical concerns.

Concerns about the use of AI in the field of nutrition can be examined under the following categories: impact on the patient, impact on the patient-dietitian relationship, impact on the dietitian, impact on the dietitian profession. One of the most frequently reported concerns with AI-based systems is the protection of patient privacy and sensitive information. In order to protect patient privacy, it is necessary to make the consent process a routine practice and not to allow unauthorized use of data[29]. The World Health Organization's 'Ethics and Governance of Artificial Intelligence for Health' guide also clearly states that for AI to be used in the healthcare field, ethical concerns and human rights must be central to the design, development and deployment of AI technologies[30]. The dietitians interviewed clearly expressed their concerns about patient privacy and data confidentiality. In the use of AI in the field of nutrition, it is necessary to pay attention to the ethical principles in the guide published by the World Health Organization.

AI's behavior towards humans is sometimes not in line with human expectations and judgments. Studies have been carried out to develop artificial empathy in order to improve the empathic appearance. However, artificial empathy can also have antisocial effects. Developing full emotional empathy involves some concerns (such as prioritizing the well-being of individuals over the well-being of society)[31]. Also, how AI will protect human values is an issue. Aligning AI with human values requires integrating value judgments such as patient rights, societal norms, and ethical guidelines with AI[32].

One of the concerns about the widespread use of AI is the possible impact of AI on human intelligence. Two theories emerge on this subject. Will AI improve human intelligence, or will over-reliance on AI technologies make people less intelligent? Nyholm seeks to answer this question in his article published in 2024[33]. The purpose of dietitians' use of AI suggests that AI can have individual effects not only on the dietitian profession but also on dietitians. In this context, it should be clearly determined where the role of AI begins and ends, and who is responsible. It should be noted that AI is an aid in this process. In this study, dietitians also expressed their questions about 'who is responsible' in the interviews.

Patient privacy and effects on the dietitian are not the only areas of concern. There are also concerns about how AI will affect the patient-dietitian relationship. While several studies have reported that AI systems used as an assistive tool could benefit the patient-healthcare professional relationship, the majority argue that an AI-led approach would harm this relationship[34]. In this study, several dietitians emphasized that AI can be an important helper, but human contact and human relations are indispensable. The inadequacy of AI in the previously mentioned areas such as empathy and human values is the basis for this view.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

One of the points emphasized by dietitians in the study is the insufficient evidence for AI. Although the number of studies estimating the potential of the use of AI in the field of nutrition is increasing, it has also been reported that AI can provide erroneous recommendations on health protection and nutritional treatment of diseases[25]. Therefore, research dietitians need to contribute to the proliferation of evidence on AI.

This study, which aims to explore dietitians' views on AI, has an important place as it is the first known study in this field. Another strength is that the sample consisted of dietitians working in different fields. Thus, insights into what role AI can play in different business areas have been obtained. As with all studies, there are some limitations in this study. The fact that the majority of the sample consists of young adults is insufficient to reflect the views of older dietitians. In addition, the fact that the interviews were held online within the limits of possibility can be considered a limitation. Although appointments are scheduled in line with the availability of dietitians, face-to-face in-depth interviews can be carried out in future studies. In addition, due to its nature, there is no generalizability of qualitative study results. Therefore, qualitative studies to be conducted in different samples will further strengthen the literature.

Conclusion

This study show that the majority of dietitians believe AI will cause a radical change in the field of nutrition. The AI's features of saving time and making routine tasks easier for dietitians have been mentioned. Despite these benefits, dietitians have some concerns about the use of AI (patient privacy, data privacy, lack of scientific evidence, etc.). It has been revealed that AI can be an aid for dietitians, but it is necessary to set boundaries well in the use of AI to protect the patient-dietitian relationship, ethical principles, and human values. In addition, continuous training programs on AI should be organized in order to break the resistance of dietitians, to benefit from AI in the most accurate way, to ensure the development of AI, and to protect clients and service quality. Increased studies evaluating the effectiveness of AI applications in the future will also eliminate concerns about insufficient evidence.

Funding: The present study was derived from a project funded by the TUBITAK 2209-A Undergraduate Research Support Program, entitled "Qualitative Analysis of Dietitians' Views on the Role of Artificial Intelligence in Nutrition" (1919B012467807).

Competing Interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

1. Azer SA, Guerrero APS. The challenges imposed by artificial intelligence: Are we ready in medical education? *BMC Med Educ.* 2023;23:680.

2. Hans D, Shevroja E, Leslie WD. Evolution in fracture risk assessment: Artificial versus augmented intelligence. *Osteoporos Int*. 2021;32(2):209-212.

3. Monteith S, Glenn T, Geddes J, Whybrow PC, Achtyes E, Bauer M. Expectations for artificial intelligence (AI) in psychiatry. *Curr Psychiatry Rep*. 2022;24(11):709-721.

4. Sullivan BA, Kausch SL, Fairchild KD. Artificial and human intelligence for early identification of neonatal sepsis. *Pediatr Res*. 2023;93(2):350-356.

5. Maheshwari K, Cywinski JB, Papay F, Khanna AK, Mathur P. Artificial intelligence for perioperative medicine: Perioperative intelligence. *Anesth Analg*. 2023;136(4):637-645.

6. Cerchia C, Lavecchia A. New avenues in artificial-intelligence-assisted drug discovery. *Drug Discov Today*. 2023;28(4):103516.

7. Fu J, Gao Q, Li S. Application of intelligent medical sensing technology. *Biosensors (Basel)*. 2023;13(8):812.

8. Pepito JA, Ito H, Betriana F, Tanioka T, Locsin RC. Intelligent humanoid robots expressing artificial humanlike empathy in nursing situations. *Nurs Philos*. 2020;21(4):e12318.

9. Horgan R, Nehme L, Abuhamad A. Artificial intelligence in obstetric ultrasound: A scoping review. *Prenat Diagn*. 2023;43(9):1176-1219.

10. Chotwanvirat P, Prachansuwan A, Sridonpai P, Kriengsinyos W. Advancements in using AI for dietary assessment based on food images: Scoping review. *J Med Internet Res*. 2024;26:e51432.

11. Ji Y, Waki K, Yamauchi T, Nangaku M, Ohe K. Using one-shot prompting of non-fine-tuned commercial artificial intelligence to assess nutrients from photographs of Japanese meals. *J Diabetes Sci Technol*. 2025:19322968241309889.

12. Lee C, Kim S, Kim J, Lim C, Jung M. Challenges of diet planning for children using artificial intelligence. *Nutr Res Pract*. 2022;16(6):801-812.

13. Haman M, Školník M, Lošťák M. AI dietician: Unveiling the accuracy of ChatGPT's nutritional estimations. *Nutrition*. 2024;119:112325.

14. Buchan ML, Goel K, Schneider CK, Steullet V, Bratton S, Basch E. National implementation of an artificial intelligence-based virtual dietitian for patients with cancer. *JCO Clin Cancer Inform*. 2024;8:e2400085.

15. Naja F, Taktouk M, Matbouli D, et al. Artificial intelligence chatbots for the nutrition management of diabetes and the metabolic syndrome. *Eur J Clin Nutr*. 2024;78(10):887-896.

16. Kenger EB, Karahan TÖ. Artificial intelligence-generated diet plans for hypertension and dyslipidemia: Adherence and nutritional insights. *Iranian Journal of Public Health*. 2025:54(6):1243.

17. Karahan TÖ, Kenger EB. ChatGPT-4o for weight management: Comparison of different diet models. *Food Science & Nutrition*. 2025:13(7):e70639.

18. Bayram HM, Arslan S, Ozturkcan A. Evaluating AI-generated meal plans for simulated diabetes profiles: A guideline-based comparison of three language models. *Journal of Evaluation in Clinical Practice*. 2025:31(7):e70295.

19. MacMillan Uribe AL, Patterson J. Are nutrition professionals ready for artificial intelligence? *J Nutr Educ Behav*. 2023;55(9):623.

20. Sosa-Holwerda A, Park OH, Albracht-Schulte K, Niraula S, Thompson L, Oldewage-Theron W. The role of artificial intelligence in nutrition research: A scoping review. *Nutrients*. 2024;16(13):2066.

21. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77–101.

22. Theodore Armand TP, Nfor KA, Kim JI, Kim HC. Applications of artificial intelligence, machine learning, and deep learning in nutrition: A systematic review. *Nutrients*. 2024;16(7):1073.

23. Atwal K. Artificial intelligence in clinical nutrition and dietetics: A brief overview of current evidence. *Nutr Clin Pract*. 2024;39(4):736-742.

24. Brown C, Story GW, Mourão-Miranda J, Baker JT. Will artificial intelligence eventually replace psychiatrists? *Br J Psychiatry*. 2021;218(3):131-134.

25. Güner E, Ülker MT. Can artificial intelligence replace dietitians? A conversation with ChatGPT. *Toros University Journal of Food, Nutrition and Gastronomy*. 2024;3(1):49-56.

26. Cleland JGF, Li C, Jones Y. Artificial intelligence needs clinical intelligence to succeed. *JACC Heart Fail*. 2020;8(7):588-591.

27. Bond A, Mccay K, Lal S. Artificial intelligence & clinical nutrition: What the future might have in store. *Clin Nutr ESPEN*. 2023;57:542-549.

28. Giansanti D. Artificial intelligence in public health: Current trends and future possibilities. *Int J Environ Res Public Health*. 2022;19(19):11907.

29. Singer P, Robinson E, Raphaeli O. The future of artificial intelligence in clinical nutrition. *Curr Opin Clin Nutr Metab Care*. 2024;27(2):200-206.

30. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. Available from: https://iris.who.int/server/api/core/bitstreams/f780d926-4ae3-42ce-a6d6-e898a5562621/content.

31. Christov-Moore L, Reggente N, Vaccaro A, et al. Preventing antisocial robots: A pathway to artificial empathy. *Sci Robot*. 2023;8(80):eabq3658.

32. Yu KH, Healey E, Leon TY, Kohane I, Manrai AK. Medical artificial intelligence and human values. *The New England Journal of Medicine*. 2024;390(20):1895–1904.

33. Nyholm S. Artificial intelligence and human enhancement: Can AI technologies make us more (artificially) intelligent? *Camb Q Healthc Ethics*. 2024;33(1):76-88.

34. Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: Some problems and solutions. *BMC Med Inform Decis Mak*. 2023;23(1):73.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 80-95.

Evaluation of Artificial Intelligence-Based Early Intervention and Education Applications[*]

Fatma GÜNGÖREN GÖKASLAN[**], Hatice KESER[***], Şenay ŞİRİN[****], Kübra ALDIRMAZ[*****]

Abstract

Aim: This study was conducted to evaluate artificial intelligence (AI)-based early intervention and educational applications developed for children from the perspective of Child Development experts, in line with the parameters of the target population, area of use, technological infrastructure, advantages, limitations, and developmental appropriateness.

Method: The study was designed as an expert-based qualitative screening study. Data were collected by examining the official websites, product promotion pages, user information pages, and privacy policies of the applications. A total of 17 early intervention and educational applications were identified through a web-based search; among these, five applications that met the predetermined criteria were included in the study. The applications reviewed were Duolingo ABC, Khanmigo, DreamBox, Suufle, and TADİS. The evaluation process was carried out by four researchers who are experts in child development, and the consistency rate among expert responses was 86.11%.

Results: The applications were evaluated across seven main parameters. The findings indicated that Duolingo ABC, Khanmigo, and DreamBox primarily supported academic skills, individualized learning, and cognitive development through adaptive mechanisms, whereas TADİS and Suufle contributed to the language, communication, and social participation skills of children with autism spectrum disorder. However, limitations were identified in certain applications regarding technological transparency, cultural adaptation, data security, and independent scientific evidence of effectiveness.

Conclusion: The results showed that AI-based early intervention and educational applications for children offer significant opportunities for personalized learning, motivation, accessibility, language development, and support for social communication skills. However, these applications should be carefully evaluated with regard to ethics, data security, scientific evidence, and developmental appropriateness. In this respect, it is recommended that AI applications developed for children be addressed not only through technological innovation but also within a holistic framework grounded in pedagogical and developmental principles.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 17.03.2026 **Kabul / Accepted:** 16.06.2026
DOI: https://doi.org/10.38079/igusabder.1911793

[*] Part of this research was presented as an oral presentation at the 6th International Congress on Child Development under the title “AI-Based Interventions Designed for Children: Early Intervention and Educational Approaches”, and only the abstract was published in the conference proceedings.

[**] PhD Student, Ankara University, Institute of Health Science, Department of Child Development, Ankara, Türkiye. E-mail: fatmagungoren456@gmail.com ORCID https://orcid.org/0000-0003-0771-4033
[***] Lecturer, Sivas Cumhuriyet University, Şarkışla Aşık Veysel Vocational School, Department of Child Care and Youth Services, Sivas, Türkiye. E-mail: haticekeser@cumhuriyet.edu.tr ORCID https://orcid.org/0000-0002-8283-7394
[****] Lecturer, Şişli Vocational School, Department of Child Care and Youth Services Department, Child, Istanbul, Türkiye. E-mail: senay.sirin@sisli.edu.tr ORCID https://orcid.org/0000-0001-9294-3855
[*****] Res. Asst., Istanbul Gelişim University, Faculty of Health Sciences, Department of Child Development, Istanbul, Türkiye. E-mail: kubralirmaz@gmail.com ORCID https://orcid.org/0000-0002-7403-5324

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Keywords: Artificial intelligence, developmental appropriateness, educational practices, early intervention.

Yapay Zekâ Temelli Erken Müdahale ve Eğitim Uygulamalarının Değerlendirilmesi

Öz

Amaç: Bu araştırma; çocuklara yönelik geliştirilen Yapay Zekâ (YZ) temelli erken müdahale ve eğitim uygulamalarını; hedef kitle, kullanım alanı, teknolojik altyapı, avantajlar, sınırlılıklar ve gelişimsel uygunluk parametreleri doğrultusunda Çocuk Gelişimi uzmanları perspektifiyle değerlendirmek amacıyla yapılmıştır.

Yöntem: Araştırma, uzman görüşlerine dayalı nitel bir tarama araştırmasıdır. Veriler, uygulamaların resmî web sayfaları, ürün tanıtım sayfaları, kullanım açıklamaları ve gizlilik politikaları incelenerek toplanmıştır. Web tabanlı taramada toplam 17 adet erken müdahale ve eğitim uygulaması listelenmiş; bunlardan belirlenen ölçütleri karşılayan beş uygulama araştırmaya dahil edilmiştir. İncelemeye Duolingo ABC, Khanmigo, DreamBox, Suufle ve TADİS uygulamaları alınmıştır. Değerlendirme süreci çocuk gelişimi alanında uzman dört araştırmacı tarafından yürütülmüş; uzman yanıtlarının tutarlılığı %86,11 olarak hesaplanmıştır.

Bulgular: Araştırma bulgularına göre uygulamalar yedi ana parametre altında değerlendirilmiştir. Bulgular; Duolingo ABC, Khanmigo ve DreamBox'ın daha çok akademik becerileri, bireyselleştirilmiş öğrenmeyi ve uyarlanabilir mekanizmalarıyla bilişsel gelişimi desteklediği; TADİS ve Suufle uygulamalarının ise otizm spektrum bozukluğu olan çocukların dil, iletişim ve sosyal katılım becerilerine katkı sağladığı belirlenmiştir. Bununla birlikte bazı uygulamalarda teknolojik şeffaflık, kültürel uyarlama, veri güvenliği ve bağımsız bilimsel etkililik kanıtları açısından sınırlılıklar saptanmıştır.

Sonuç: Araştırma sonucunda, çocuklara yönelik AI temelli ve erken müdahale ve eğitim uygulamalarının kişiselleştirilmiş öğrenme, motivasyon, erişilebilirlik, dil gelişimi ve sosyal iletişim becerilerinin desteklenmesi açısından önemli fırsatlar sunduğu; ancak etik, veri güvenliği, bilimsel kanıt ve gelişimsel uygunluk boyutlarında dikkatle değerlendirilmesi gerektiği saptanmıştır. Bu doğrultuda, çocuklara yönelik yapay zekâ uygulamalarının yalnızca teknolojik yenilik açısından değil, pedagojik ve gelişimsel ilkeler çerçevesinde bütüncül olarak ele alınması önerilmektedir.

Anahtar Sözcükler: Yapay zekâ, gelişimsel uygunluk, eğitim uygulamaları, erken müdahale.

Introduction

The education sector is rapidly transforming due to AI, which can support children's cognitive and language development and personalize learning[1,2]. These applications are also beneficial for children with learning difficulties, though most early childhood AI tools are still designed for older learners[3]. Research has largely focused on technology's impact on teachers or higher education, leaving a gap in studies on AI interventions that directly enhance children's cognitive, emotional, and social development[4]. In Türkiye, local applications like TADIS and Suufle remain underexplored. AI interventions use technologies such as speech recognition, natural language processing, emotion analysis, social simulation, and intelligent tutoring[5,6]. Since the 1990s, projects like Carnegie Mellon's Project LISTEN/RoboTutor and Justine Cassell's Embodied Conversational Agents[7] have supported reading and social skills, while BabyX[8] (2013) and Squirrel AI Learning[9] (post-2014) demonstrated interactive avatars and adaptive personalized learning. Between 2023 and 2025, generative AI and avatar technologies in early childhood education accelerated, focusing on smart tutors, operational tools, digital assistants, and automated content generation, transforming both individual learning and the broader education ecosystem[10]. AI-based technologies have both benefits and risks for child development. They support cognitive and language skills through tools like audio reading and pronunciation systems[11,12] and improve learning for children with attention deficits via personalized content[13]. AI can also enhance problem-solving and critical thinking[1,6]. Socially, robots and avatars support interaction skills, especially in children with Autism Spectrum Disorder (ASD)[7,14,15], though excessive use may increase isolation in nexurotypical children[16]. Creative AI tools foster digital literacy and early interest in STEM[6], but risks include bias, emotional overattachment, and confusion between virtual and real experiences[13,16]. Additionally, robotics and Augmented Reality (AR) applications contribute to both gross and fine motor skill development. Programmable toys, such as LEGO Mindstorms and Ozobot, can particularly develop children's physical coordination and logical sequencing skills[17]. AI offers more responsive, personalized, and motivating learning for children, particularly in early childhood and primary education, when cognitive and social development is critical. In education, AI can function as a tool or be integrated with diverse pedagogical approaches to create flexible, student-centered, and customizable learning experiences. With limited research on this topic, this study evaluates both international and local AI-based early intervention and educational applications for children, examining target audiences, technological infrastructure, benefits and drawbacks, data security, and developmental appropriateness. The study highlights the current state of AI in early childhood education and provides a conceptual framework for child-centered digital learning solutions.

Material and Methods

Study Design and Data Collection

This study is an expert-based qualitative review that evaluates AI-based early intervention and educational applications for children from a pedagogical perspective. The analysis relied on publicly available documents (e.g., websites, manuals, privacy policies) rather than experimental use. The process included parameter selection, individual evaluations, and expert consensus. Of the 17 applications identified through web searches, only five met the inclusion criteria and were included in the final sample.

The applications included in the review were determined according to the following criteria:

1. The application explicitly targets children as its target audience or states that it is suitable for children's use (n=10),

2. Having an educational, developmental, or communicative purpose (n=13),

3. Incorporating an AI-based function (e.g., personalized learning, natural language processing, automated feedback, adaptive content delivery, etc.) (n=16),

4. Possession of accessible digital documentation at a level that allows for evaluation (website, product pages, privacy policies, and/or open access reports) (n=17).

The applications 'Duolingo ABC', 'Khanmigo', 'DreamBox', 'Suufle', and 'TADIS (Turkish Alternative and Augmentative Communication System)', which meet all of these criteria, were included in the study.

Determination of Evaluation Parameters

The evaluation parameters were developed inductively, based on expert input, rather than using a pre-existing scale. They were derived from a preliminary analysis of the applications' websites, technical documentation, and user interfaces. Recurring themes were grouped into a set of parameters organized around seven main categories: Target Audience, Area of Use, Technological Infrastructure, Advantages, Limitations, Data Security, and Developmental Suitability.

Validity and Reliability

To determine the reliability of the assessment findings, the consistency of expert responses was calculated. The assessments were compared at the application×parameter level (6×6=36 cells); using the Miles and Huberman (1994)[18] formula, the inter-expert agreement was calculated as 86.11%. This ratio is consistent with the threshold of 80% and above, which is often used for inter-coder reliability in qualitative studies and is considered "acceptable/high agreement" indicating that the expert assessments are consistent and reliable[18].

Ethical Statement

Since the study relied solely on publicly available web-based data, no human participants were involved, and no consent procedures were needed. The findings do not demonstrate real-world effectiveness but offer a pedagogical framework based on official information and expert evaluations.

Data Analysis

The evaluation was conducted by four Child Development experts, who reviewed each application's online documentation and produced qualitative assessments against predefined parameters. Their evaluations were compared to identify common and differing points; shared views were synthesized, while discrepancies were discussed until consensus was reached based on available evidence. Cases with insufficient information were also noted.

Results

AI-supported applications aimed at, suitable for children's use, or suitable for children were examined. Each application is analyzed below within the framework of the defined themes and presented in detail.

Table 1 below presents the evaluation of the Duolingo ABC application based on the defined parameters. Table 1 indicates that the Duolingo ABC application is an individualized, interactive language-learning tool suitable for children in early childhood.

Table 1. Results of the Duolingo ABC Application

Program Name	Duolingo ABC
Target Audience	Aimed at preschool and first-grade elementary school students
Application Area	Aims to improve children's comprehension skills by increasing their phonological awareness, vocabulary, and reading fluency. The program includes phonetic exercises that teach letter sounds, syllable types, and spelling patterns, as well as fluency exercises that support fast, error-free reading.
Technological Infrastructure	The technologies included in the application are voice recognition, which aims to develop correct reading habits by evaluating pronunciation; interactive learning, which includes interactive activities such as drag-and-drop, letter writing, and audio reading; and offline access, which allows use even without an internet connection. It includes AI-powered personalization to achieve faster progress.
Advantages	The app is free and ad-free. Parents and teachers can monitor children's progress and intervene when necessary.
Limitations	The app offers support in English, German, Spanish, French, Portuguese, Japanese, and Korean. This may limit accessibility for children whose native language is Turkish or other languages. The content offered has limitations for advanced reading and writing skills. (Data is valid for 2025; additional language options may be added in subsequent years.)
Data security	It has been stated that all necessary technical measures have been taken to protect personal data and ensure information security. However, in line with the principle of "impossibility of absolute protection" in the digital security literature, it has been clearly stated that data may be accessible in the event of a cyberattack with sufficient resources and determination.
Developmental Suitability	The US-developed Duolingo ABC app has the potential to reduce learning inequalities through its adaptive and personalized approach for different skill levels. The effectiveness of AI-supported language learning depends on usage frequency, individual learning strategies, and motivation. Developmentally, the app supports cognitive growth (academic skills, reading comprehension), language development (language learning, reading fluency), and social-emotional development through personalized and interactive activities.

The following table presents the evaluation of Khanmigo based on the specified parameters. The Khanmigo application, detailed in Table 2, has been evaluated as an academic guidance tool that provides content in various subjects for teachers, students, and parents.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Table 2. Results of the Khanmigo App

Program Name	**Khanmigo**
Target Audience	It is aimed at students, teachers, and parents from first grade through all levels of primary school.
Application Area	This AI-based educational system supports students, teachers, and parents. For students, it provides guidance in subjects such as math, science, social studies, coding, financial literacy, and humanities, encourages problem-solving through the Socratic method, and offers feedback on homework and exams. For teachers, it assists with lesson planning, exam creation, student grouping, progress monitoring, material preparation, and content generation. For parents, it enables viewing chat history, receiving safety alerts, and monitoring their children's learning.
Technological Infrastructure	It is based on OpenAI's GPT-4 model. It works in conjunction with Khan Academy's content library
Advantages	The application offers free services to teachers in regions where it is available. It can be used anytime and anywhere.
Limitations	The application is paid for parents (monthly $4, yearly $44/2025). It only provides services in English. Student access is limited to parental or school district approval. The number of children a parent can access through the system is limited to 10.
Data Security	The platform contains no third-party advertisements and implements restrictions for users under 13, including automatic controls to prevent the sharing of personal information. Parental consent is required for account creation and data sharing, and parents have full access to manage, edit, or delete their children's accounts.
Developmental Appropriateness	The US-developed Khanmigo app uses a Socratic questioning approach, promoting metacognitive skills by encouraging students to think rather than giving direct answers. It functions as a cognitive guide, not just a knowledge-transfer tool. Limited English-language support may restrict developmental inclusiveness, and teacher guidance remains important to address potential mismatches with students' readiness levels. Developmentally, the app supports cognitive growth through academic content and Socratic thinking, and social-emotional development through feedback and fostering independent problem-solving.

The following table presents the evaluation of DreamBox based on the specified parameters. As detailed in Table 3, the DreamBox application is an interactive, individualized educational tool for mathematics and reading skills for students from preschool through 8th grade, offering teachers and parents the opportunity to monitor student progress and intervene.

Table 3. Results of the DreamBox Application

Program Name	**DreamBox**
Target Audience	Suitable for use by preschoolers, students in grades 1–8, teachers, parents, and educational institutions
Application Area	Content is provided for math topics such as numbers and operations, fractions, ratios, algebraic thinking, and geometry; as well as reading skills such as reading fluency, vocabulary, reading comprehension, and word analysis.
Technological Infrastructure	DreamBox utilizes its proprietary Intelligent Adaptive Learning™ (IAL) technology. This technology analyzes students' interactions in real time and adapts lesson content and difficulty levels to individual needs.
Advantages	The personalized learning feature provides lesson modules tailored to students' individual needs. Parents and teachers can monitor students' progress and intervene.
Limitations	The application is available to families (monthly $19.95/2025) and individual subscribers (monthly $12.95/2025). It is currently only available in English. Student access is limited to approval by parents or school districts.
Data Security	The DreamBox team has committed to protecting personal data. It has stated that it ensures data security, does not share service content with any unrelated companies, and that accessed and shared information is encrypted during transmission and stored in encrypted form.
Developmental Appropriateness	The US-developed DreamBox app provides interactive, gamified learning tailored to individual differences in attention, perception, and memory. It particularly supports the development of mathematical skills while offering an enjoyable, personalized experience. As AI may not fully assess cognitive needs, parental and teacher guidance is recommended. Developmentally, the app supports cognitive growth through math, geometry, and reading, and social-emotional development via personalized education that responds to student interactions.

The following table presents the evaluation of the Suufle application based on the specified parameters. The Suufle application has been shown to be a scientifically based, customizable, and developmentally targeted intervention tool. However, it has been found to have limitations in accessibility, user competence, and technological transparency.

Table 4. Results of the Suufle Application

Program Name	**Suufle**
Target Audience	Developed for children with autism spectrum disorder.
Application Area	Used to support the language and communication skills of children with autism spectrum disorder.
Technological Infrastructure	Suufle is supported by Harvard Innovation Lab, one of the world's most prestigious entrepreneurship centers. However, no information regarding their technological infrastructure has been disclosed.
Advantages	The application is an innovative educational tool designed to effectively develop language and communication skills in individuals with autism, grounded in scientific principles. Parents and professionals can begin using the application after receiving training
Limitations	The Suffle app costs 6,000 TL for individual users and parents, including unlimited use on a single device, and 30,000 TL for institutions, including unlimited use on 10 devices. The app's usage fees are valid for 2025. It can be considered a limitation that researchers review the app based on the information provided and are not users of the app.
Data Security	The application indicates that personal data is protected against unauthorized access, alteration, or disclosure. However, it acknowledges that, due to the nature of Internet-based transmission, complete data security cannot be guaranteed, highlighting both protective measures and inherent limitations in digital environments.
Developmental Appropriateness	The Turkey-developed Suufle app targets social interaction and communication skills, key areas of difficulty for children with ASD. Its personalized approach supports individual developmental needs and provides content aimed at language development, cognitive skills, and social communication, offering potentially reliable, measurable, and effective learning opportunities.

The following table presents the evaluation of the TADİS application based on the specified parameters. TADİS is an alternative and augmentative communication system designed for children who lack verbal communication or require support with communication. It operates on iOS, Android, and web platforms, supporting language and communication development through eye-controlled technology, voice-enabled visuals, and customizable content. However, it is noteworthy that the application requires paid certification training and has certain usage limitations.

Table 5. Results of the Turkish Alternative and Augmentative Communication System (TADİS) Application

Program Name	**TADİS (Turkish Alternative and Augmentative Communication System)**
Target Audience	Developed for children who cannot communicate verbally and need support in verbal communication. It aims to remove communication barriers by offering technological solutions for individuals with diagnoses such as autism, cerebral palsy, and Down syndrome, or those who experience speech and communication difficulties.
Application Area	It is used to provide alternative communication for children who cannot speak and to support verbal communication for children with speech difficulties.
Technological Infrastructure	The TADİS application is available on iOS, Android, and web platforms. It includes Eye Control technology, specifically designed for individuals who cannot use their hands effectively.
Advantages	It is possible to prepare content appropriate to the developmental level of every child experiencing communication difficulties, regardless of diagnosis. TADİS contains 2000 voiced images.
Limitations	To use TADİS, therapists or families must complete a practitioner certification training program worth 11,000 TL. Parents may be limited in their ability to individualize and step-by-step target behaviors. Another limitation is that researchers may only examine the information provided about the application and cannot be its users.
Data Security	The application processes personal data in accordance with Turkey's KVKK (Personal Data Protection Law), collecting information such as name, email, phone number, IP address, device details, and usage history. Data is not shared with third parties for commercial purposes and is protected in accordance with ISO 27001 standards. It is stored while the account is active and deleted within 30 days upon request. Parental or guardian consent is required for users aged 13–18.
Developmental Appropriateness	The Turkey-developed TADİS app is a developmentally appropriate alternative and supportive communication system that aids children's language and communication growth. It allows children to express feelings, desires, and thoughts using images and symbols, expand vocabulary, and gradually build sentences. Therapists or families can add images and record voices, with options for different audio outputs, enabling personalized communication support. The app's step-by-step progression—from single-symbol selections to longer sentences—and its "Press-to-speak" feature enhance expressive language skills, encourage functional communication, and support social-emotional development by fostering social interaction, self-confidence, and motivation. Additionally, using symbols to make choices and construct meaning promotes cognitive development and communication-based learning.

Discussion

In this study, five AI-based applications developed to support children's developmental areas were evaluated by four experts using the following parameters: usage area, target audience, developmental appropriateness, data security, advantages, and limitations. It was observed that the applications cater to different needs and user groups. Duolingo, DreamBox, and Khanmigo[19–21] support academic skills such as math, reading, problem solving, language, and conceptual learning in typically developing children, while Suufle and TADİS[22,23] provide augmentative communication systems for children with autism spectrum disorder, supporting basic language and social communication skills. Language development occurs based on experience, the child's learning strategies, needs, and living conditions[24]. Noam Chomsky argues that children can acquire language with minimal environmental input[25]. In this sense, Duolingo supports natural and enjoyable learning through short, repetitive activities suited to children's limited attention spans, which involve noticing, selecting, and sustaining focus on relevant stimuli[26]. Regular use of the Duolingo[19] application in foreign/second language learning has been found to be effective in language learning, receptive language skills, interest level, vocabulary knowledge[27]; receptive and expressive language skills[28]; and the retention of learned words in vocabulary and long-term memory[29]. In practice, the fact that students who complete short but frequent lessons make more progress, especially with regular use, supports the idea that technology-based language-learning applications can develop various language skills[28,30]. However, the application alone may not be sufficient as a comprehensive language learning tool. This is consistent with views that learning occurs through social interaction in the early years[31]. Scientific research reports on the Duolingo ABC[19] application are limited. According to an independent research report, kindergarten and first-grade students who used the app showed significant improvements in phonological awareness, letter recognition, and word reading skills. It was stated that children who initially had lower reading skills than others made more progress and that meaningful gains were achieved with an average daily use of 10 minutes[32]. Regular, continuous use of the application can reduce learning inequalities through its adaptability to different learning levels and personalization. The effectiveness of AI-supported language learning depends on usage frequency, individual learning strategies, and motivation. Their success relies not only on technological features but also on the structure of the learning process, and they are more developmentally effective when combined with adult support and other learning activities. While Duolingo's[19] natural, fun, and repetition-based learning approach creates a suitable foundation for the early years, it is seen that educational assistants such as Khanmigo[21] structure the learning process in a more conscious way, along with metacognitive strategies and self-regulation skills that develop in middle childhood. Metacognitive learning processes involve the use of knowledge the reader possesses and retains during the mental learning strategy[33]. Metacognitive and self-regulation skills emerge in middle childhood (ages 6–12) as thinking speed and neural connectivity increase. Features like reminders to parents, gamification, and achievement badges support motivation[34,35]. Khanmigo showed that 30 minutes of weekly use improved exam scores[34], using the app three or more days per week increased math scores[35,36] and teachers enhanced lessons by adapting content to students' interests[37]. It was determined that teachers saved time through planning and differentiation[38]. Students are directed to Khanmigo before asking their teachers for support, students who use the application effectively are recognized, and thus productive help-seeking behavior is encouraged[39]. When evaluated from a developmental perspective, Khanmigo[21] as an educational assistant stands out for providing a learning environment compatible with students' cognitive and academic development characteristics. DreamBox[20], another application targeting a similar age group, aims to develop academic skills, particularly in mathematics and reading, with its adaptable structure that takes individual differences into account. It is known that developmental characteristics and developmental speed are unique to each individual due to differences in each child's genetic characteristics and environmental opportunities[40]. DreamBox[20] is a tool developed for school-age children. The cognitive processes of school-age children are structured to select and transform information. There is a linear relationship between the time spent on the DreamBox[20] math application and success[41]. Students who complete more math lessons achieve higher scores in end-of-year assessments[42], students who completed at least 55 minutes of lessons per week had higher year-end math achievement scores[43], and students who completed at least 100 DreamBox[20] reading lessons were more successful in reading proficiency compared to those who did not complete them. Teachers using the DreamBox[20] application reported that their students asked thought-provoking questions; parents reported that their children actively participated and focused on learning at school· In summary, it was determined that as students' reading skills improved, their interest in reading and self-confidence also increased. Accordingly, it can be said that the DreamBox[20] reading application is an effective reading development tool not only for highly motivated students but also for those with low motivation; as reading skills improve, students' self-confidence and interest in reading also increase significantly[44]. AI-based applications developed in Türkiye primarily target children with autism spectrum disorder and other diagnostic groups, supporting language, communication, and social skills through direct use[22,23]. Research shows that well-designed digital applications can positively influence language development[45,46], especially when content is reliable and parents are involved. Suufle[22] claims to support language development with gamified, research-based content, but independent scientific evidence is lacking. The website provides "before and after" skill comparisons, yet details on methodology, sample size, usage duration, and practitioner involvement are unavailable. Bektaş and Görgü (2025)[47] evaluated TADİS[23] for children with moderate to severe intellectual disabilities and autism. It was stated that the final test scores of the participants in the experimental group were significantly higher than those in the control group. This result shows that the TADİS[23] program is an effective tool for improving the language skills of individuals with special needs. Additionally, it should be noted that scientific research on TADİS[23] is limited, as in the case with Suufle[22]. The effectiveness of the AI-based applications reviewed[19,23] aligns with recent literature, which shows positive impacts on personalized learning and motivation but is limited by methodological diversity, short follow-ups, and lack of international samples[48]. Observed benefits, such as Duolingo ABC[19] and DreamBox[20] improving reading and math, reflect these trends. However, international reports stress that effectiveness depends not only on algorithms but also on contextual factors such as equitable access, cultural adaptation, teacher training, and data security[2,49]. Accordingly, the challenges identified in this study—limited adaptation of international applications to Turkish content and scarce research on local applications—mirror broader critiques, suggesting that laboratory- or company-reported successes may not fully translate into real-world settings. From an ethical and child-centered design perspective, UNICEF guidelines recommend evaluating AI applications in line with children's rights, including child protection, transparent algorithms, data minimization, and child participation[50]. This study found limited reporting on such practices, especially for local applications[22,23], hindering their scalability and institutional endorsement. Ethical compliance and transparency should therefore be key evaluation criteria. Literature from 2022–2024 also highlights concerns about bias and inclusivity in AI education[51], yet program reports often neglect these risks, which is particularly critical for applications targeting children with special needs, as misclassification or inappropriate recommendations can reduce effectiveness and compromise safety[22,23]. This study highlights the potential of AI-based educational and support applications[19,23] in early intervention and learning, based on expert evaluations. These systems offer personalized learning, motivation, and accessibility: Duolingo ABC and DreamBox support cognitive development, while TADİS and Suufle enhance communication and social participation. However, major gaps exist in data transparency, cultural adaptation, and scientific evidence of effectiveness, aligning with reports that emphasize the importance of data privacy, ethical evaluation, and algorithmic risk for sustainable pedagogical benefits[2,50,51]. Given their focus on children, ethical practices and data security are especially critical, and the handling of user data remains a key concern.

Limitations

As the AI-based applications in this study were not tested with children, evaluations of effectiveness and pedagogical benefits rely on official website information and available scientific evidence rather than actual usage. Limited transparency of web documents and scarce independent research, especially for domestic applications, constrain generalizability. Nonetheless, the study provides a systematic evaluation framework that compares AI applications for children using multidimensional criteria and assessments from four experts, though the sample is limited to applications identified via specific keywords.

Conclusions and Recommendations

The AI-based applications in this study were evaluated using official websites, developer-provided information, and published research. Limited independent studies on domestic applications restricted the analysis, and the lack of direct testing with children limits validation of real-world effectiveness. Overall, AI applications for children should be considered not only as technological tools but also in terms of ethical, developmental, and educational suitability.

REFERENCES

1. Holmes W, Bialik M, Fadel C. *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston, MA: Center for Curriculum Redesign; 2019.

2. UNESCO. *AI and Education: Guidance for Policy-Makers*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386012. Accessed January 7, 2026.

3. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education: where are the educators? *Int J Educ Technol High Educ*. 2019;16(1):1-27.

4. OECD. *AI and the Future of Skills, Volume 1: Capabilities and Assessments*. OECD Publishing; 2021. doi: 10.1787/5ee71f34-en.

5. Woolf N, Quinn J. Students' perceptions of instructional design implementation in a contextualized learning activity. *Educ Technol Res Dev*. 2009;57(1):25-43.

6. Luckin R, Holmes W, Griffiths M, Forcier LB. *Intelligence Unleashed: An Argument for AI in Education*. London: Pearson; 2016.

7. Cassell J. Embodied conversational agents: representation and intelligence in user interfaces. *AI Mag*. 2000;22(4):67-83. doi: 10.1609/aimag.v22i4.1593.

8. Sagar M. *BabyX Project*. Commercial project. Soul Machines; 2013.

9. Nye BD, Graesser AC, Hu X. AutoTutor and family: a review of 17 years of natural language tutoring. *Int J Artif Intell Educ*. 2014;24(4):427-469.

10. Tartaro A, Cassell J. Playing with virtual peers: bootstrapping contingent discourse in children with autism. In: *Proceedings of the 8th International Conference on Intelligent Virtual Agents (IVA 2008)*. Springer; 2008:294-300. doi: 10.1007/978-3-540-85483-8_30.

11. Mostow J, Aist G, Burkhead P, et al. Evaluation of an automated reading tutor that listens: comparison to human tutoring and classroom instruction. *J Educ Comput Res*. 2003;29(1):61-117. doi: 10.2190/06AX-QW99-EQ5G-RDCF.

12. Farhood H, Nyden M, Beheshti A, Muller S. Artificial intelligence-based personalised learning in education: a systematic literature review. *Discover Artificial Intelligence*. 2025;5(1):331. doi: 10.1007/s44163-025-00598-x

13. Zero To Three. *Using Technology to Support Young Children's Learning*. Zero To Three. https://www.zerotothree.org/resource/using-technology-to-support-young-childrens-learning. Accessed December 24, 2025.

14. Qin L, Wang H, Ning W, Cui M, Wang Q. New advances in the diagnosis and treatment of autism spectrum disorders. *Eur J Med Res*. 2024;29(1):322.

15. Lai T, Xie C, Ruan M, Wang Z, Lu H, Fu S. Influence of artificial intelligence in education on adolescents' social adaptability: the mediatory role of social support. *PLoS One*. 2023;18(3). doi: 10.1371/journal.pone.0283170.

16. UNICEF. *Policy Guidance on AI for Children*. UNICEF. https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children. Accessed January 2026.

17. Resnick M, Maloney J, Monroy-Hernández A, et al. Scratch: programming for all. *Commun ACM*. 2009;52(11):60-67. doi: 10.1145/1592761.1592779.

18. Miles MB, Huberman AM. *Qualitative Data Analysis: An Expanded Sourcebook*. 2nd ed. SAGE Publications; 1994.

19. Duolingo ABC. https://abc.duolingo.com/. Accessed July 19, 2025.

20. DreamBox. https://www.dreambox.com/. Accessed July 19, 2025.

21. Khanmigo. https://www.khanmigo.ai/. Accessed July 19, 2025.

22. Suufle. https://suufle.com/. Accessed August 10, 2025.

23. TADİS. https://www.TADİS.com.tr/. Accessed August 10, 2025.

24. Korkmaz B. *Dil ve Beyin: Çocuklarda Dil ve Konuşma Bozuklukları*. Yüce Yayınları; 2005:3.

25. Güngören F. Chomsky. In: Özyürek A, ed. *Kuramcıların Perspektifinden Gelişim Kuramlarına Bir Bakış*. Nobel; 2024:171.

26. Şahin S. Bilişsel gelişim. In: Metin NB, ed. *Doğum Öncesi Dönemden Ergenliğe Çocuk Gelişimi*. Pegem; 2021:91.

27. González-Fernández B, de la Viña I. The effectiveness of app-based and classroom-based instruction on L2 learning and motivation. *Lang Learn Technol*. 2025;29(1):1-18.

28. Smith B, Jiang X, Peters R. The effectiveness of Duolingo in developing receptive and productive language knowledge and proficiency. *Lang Learn Technol*. 2024;28(1).

29. Kittredge AK, Hopman EW, Reuveni B, Dionne D, Freeman C, Jiang X. Mobile language app learners' self-efficacy increases after using generative AI. *Front Educ*. 2025;10:1499497.

30. Sudina E, Plonsky L. The effects of frequency, duration, and intensity on L2 learning through Duolingo: a natural experiment. *J Second Lang Stud.* 2024;7(1):1-43.

31. Şirin Ş. Bandura. In: Özyürek A, ed. *Kuramcıların Perspektifinden Gelişim Kuramlarına Bir Bakış*. Nobel; 2024:81.

32. Channa MA, Nordin ZS, Siming IA, Chandio AA, Koondher MA. Developing reading comprehension through metacognitive strategies: a review of previous studies. *Engl Lang Teach.* 2015;8(8):181-186.

33. Berk L. *Bebekler ve Çocuklar*. Nobel. 2020:442-443.

34. Oreopoulos P, Gibbs C, Jensen M, Price J. *Teaching Teachers to Use Computer Assisted Learning Effectively: Experimental and Quasi-Experimental Evidence*. National Bureau of Economic Research; 2024. Working Paper 32388. doi: 10.3386/w32388.

35. Ferman B, Finamor L, Lima L. *Are Public Schools Ready to Integrate Math Classes With Khan Academy?* University Library of Munich; 2019.

36. Khan Academy. *How Oasis Elementary South Scaled Math Mastery With Khan Academy Districts*. https://www.kisa.link/XAqSs. Accessed February 14, 2026.

37. Khan Academy. *Newark Public Schools*. https://www.kisa.link/SYZYo. Accessed February 14, 2026.

38. Khan Academy. *Cold Spring*. https://www.kisa.link/pbnpL. Accessed February 14, 2026.

39. Khan Academy. *How Oasis Elementary South Scaled Math Mastery With Khan Academy Districts*. https://www.kisa.link/XAqSs. Accessed February 14, 2026.

40. Doğan Ö, Şengül EA. Büyüme ve gelişme. In: Metin NB, ed. *Doğum Öncesi Dönemden Ergenliğe Çocuk Gelişimi*. Pegem; 2021:8.

41. Harvard University, Center for Education Policy Research. *DreamBox Learning Achievement Growth in the Howard County Public School System and Rocketship Education*. https://www.kisa.link/gDATs.

42. LearnPlatform. *DreamBox Math Drives Higher Achievement for K-8 Students in Large Arizona District*. https://www.kisa.link/YXFtZ. Accessed August 18, 2025.

43. LearnPlatform by Instructure. *DreamBox Math (2021-22)*. https://www.kisa.link/HPGwa. Accessed February 13, 2026.

44. Discovery Education. *Reading Motivation and Reading Success: A Two-Way Street*. https://www.kisa.link/OvHvg. Accessed August 18, 2025.

45. Neumann MM, Neumann DL. The use of touch-screen tablets at home and pre-school to foster emergent literacy. *J Early Child Lit*. 2017;17(2):203-220.

46. Kucirkova N, Sheehy K, Messer D, Flewitt R. *Digital Literacy in Early Childhood Education: Theoretical Perspectives and Practical Applications*. Routledge; 2020.

47. Bektaş E, Görgü E. Türk alternatif iletişim destek sistemlerinin engelli bireylerin dil gelişimi üzerindeki etkisi. *J Civ Educ Res*. 2025;9(2):55-65.

48. Wang S. Artificial intelligence in education: a systematic literature review. *Comput Educ*. 2024;198:104804. doi: 10.1016/j.compedu.2024.104804.

49. OECD. *The Potential Impact of Artificial Intelligence on Equity and Inclusion in Education*. OECD Publishing; 2024. doi: 10.1787/15df715b-en.

50. UNICEF Innocenti. *Policy Guidance on Artificial Intelligence for Children*. UNICEF Innocenti; 2021.

51. Personal Data Protection Authority (KVKK). Law No. 6698 on the Protection of Personal Data. *Official Gazette*. January 7, 2016.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 96-108.

Unsupervised Multivariate Analysis of Gait Characteristics and Body Composition in Physiotherapy Students

Başar ÖZTÜRK*, Nadide Gizem TARAKÇI FİLİZ**, Berkay Zülfikar KIZILIRMAK***, Muazzez GARİPAĞAOĞLU****

Abstract

Aim: Measurements of locomotion performance (through quantitative gait analysis) and of body composition offer detailed information on locomotor and biomechanical functions, respectively. The correlation between gait features and body composition among physiotherapy students has yet been extensively investigated using multivariate analyses. We thus evaluated the relationships between body composition components and the parameters of the walk in physiotherapy students, through an unsupervised exploratory approach.

Methods: This cross-sectional study included 48 physiotherapy students (35 women and 13 men, aged mean (SD): 21.3 (1.8) years). Measurements of body composition were done using bioelectrical impedance and consisted in body mass index (BMI), body fat percentage, muscle percentage, total body water percentage, and resting energy expenditure. Measurements of gait were performed with the OptoGait system, comprising cadence, contact time, and propulsive phase. Data was analysed using Pearson's correlation and with both hierarchical clustering and principal component analysis (PCA) in order to find out multimodal relationships between gait and body composition variables.

Results: The results of Hierarchical cluster analyses yielded two broad functional domains built on gait and body composition variables. A first domain encompassing the number of steps, proportion of propulsion, and % of muscle, and total % of water was conceptualized as a performance-based group. A second domain with % body fat, BMI and contact time was thought to represent a mass-based group linked with less desirable gait features. % total body water was seen to be a bridging element between locomotory performance and body mass variables.

A positive association was seen between % muscle mass and the number of steps ($r=0.62$, $p<0.001$) while % body fat was negatively associated with this latter variable ($r= -0.54$, $p<0.001$). In principal components analyses, the first two components accounted for 73.6 % of the total variance.

Conclusion: Body composition metrics were strongly related to gait patterns in apparently healthy physiotherapy students, with a greater proportion of muscle mass and total body water related to a higher frequency and a more propulsive phase of gait, whilst increased body fat proportion and BMI were correlated with a prolonged contact time and a lower frequency of gait. Further exploratory multivariate analyses would seem appropriate to expand on the biomechanical relationships of young adults.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 24.11.2025 **Kabul / Accepted:** 21.07.2026
DOI: https://doi.org/10.38079/igusabder.1828161
* Assoc. Prof., Fenerbahçe University, Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Istanbul, Türkiye. E-mail: fztbasarozturk@gmail.com ORCID https://orcid.org/0000-0001-5763-6008
** Asst. Prof., Fenerbahçe University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: nadide.tarakci@fbu.edu.tr ORCID https://orcid.org/0000-0001-5449-7866
*** Res. Asst., Fenerbahçe University, Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Istanbul, Türkiye. E-mail: berkay.kizilirmak@fbu.edu.tr ORCID https://orcid.org/0009-0000-9493-4549
**** Prof. Dr., Fenerbahçe University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: muazzez.garipagaoglu@fbu.edu.tr ORCID https://orcid.org/0000-0002-5323-7154

ETHICAL STATEMENT: The study was approved by the Fenerbahçe University Ethics Committee, Türkiye (Decision No: 2025/041, Date: April 2025). All participants provided written informed consent prior to data collection.

Keywords: Gait analysis, body composition, biomechanics, physiotherapy students, principal component analysis, hierarchical clustering.

Fizyoterapi Öğrencilerinde Yürüme Karakteristikleri ve Vücut Kompozisyonunun Çok Değişkenli Analizi

Öz

Amaç: Lokomotor performansın nicel yürüme analizi ile, vücut kompozisyonunun ise ilgili ölçüm yöntemleriyle değerlendirilmesi; bireylerin hareket ve biyomekanik özellikleri hakkında ayrıntılı bilgi sağlamaktadır. Fizyoterapi öğrencilerinde yürüme özellikleri ile vücut kompozisyonu arasındaki ilişki çok değişkenli analiz yöntemleri kullanılarak henüz yeterince araştırılmamıştır. Bu nedenle, bu çalışmada fizyoterapi öğrencilerinde vücut kompozisyonu bileşenleri ile yürüme parametreleri arasındaki ilişkiler denetimsiz keşifsel bir yaklaşım kullanılarak incelenmiştir.

Yöntem: Bu kesitsel çalışmaya 48 fizyoterapi öğrencisi (35 kadın, 13 erkek; yaş ortalaması ± standart sapma: 21,3 ± 1,8 yıl) dahil edildi. Vücut kompozisyonu ölçümleri biyolojik empedans analizi kullanılarak gerçekleştirildi ve vücut kitle indeksi (VKİ), vücut yağ yüzdesi, kas yüzdesi, toplam vücut suyu yüzdesi ve bazal metabolizma hızını kapsadı. Yürüme değerlendirmesi OptoGait sistemi ile yapıldı ve kadans, temas süresi ile itici faz parametreleri kaydedildi. Veriler Pearson korelasyon analizi, hiyerarşik kümeleme ve temel bileşen analizi (PCA) kullanılarak analiz edilerek yürüme ve vücut kompozisyonu değişkenleri arasındaki çok boyutlu ilişkiler araştırıldı.

Bulgular: Hiyerarşik kümeleme analizi sonucunda yürüme ve vücut kompozisyonu değişkenlerinden oluşan iki temel fonksiyonel alan belirlendi. İlk kümede kadans, itici faz, kas yüzdesi ve toplam vücut suyu yüzdesi yer alırken, bu küme lokomotor performans ile vücut kompozisyonu arasındaki ilişkiyi yansıtmaktadır. İkinci küme, vücut yağ yüzdesi ve BKİ gibi vücut kompozisyonu değişkenlerinden oluşmuş olup, daha uzun temas süresi ve daha düşük yürüme performansı ile ilişkilidir. Toplam vücut suyu yüzdesi ise yürüme bileşenleri ile vücut kompozisyonu arasında geçiş değişkeni olarak öne çıkmıştır. Kas yüzdesi ile kadans arasında pozitif ($r=0{,}62$; $p<0{,}001$), vücut yağ yüzdesi ile kadans arasında ise negatif korelasyon ($r=-0{,}54$; $p<0{,}001$) saptanmıştır. Temel bileşen analizinde ilk iki bileşen toplam varyansın %73,6'sını açıklamıştır.

Sonuç: Sağlıklı fizyoterapi öğrencilerinde vücut kompozisyonu değişkenleri ile yürüme parametreleri arasında anlamlı ilişkiler bulunmaktadır. Daha yüksek kas yüzdesi ve toplam vücut suyu yüzdesi, daha yüksek kadans ve daha uzun itici faz ile ilişkiliyken, daha yüksek vücut yağ yüzdesi ve VKİ değerleri daha uzun temas süresi ve daha düşük kadans ile ilişkilidir. Çok değişkenli keşifsel analiz yöntemleri, fizyoterapi öğrencilerindeki biyomekanik ilişkilerin daha kapsamlı biçimde yorumlanmasına katkı sağlayabilir.

Anahtar Sözcükler: Yürüme analizi, vücut kompozisyonu, biyomekanik, fizyoterapi öğrencileri, temel bileşen analizi, hiyerarşik kümeleme.

Introduction

Gait analysis is a core component of evaluating an individual's locomotion. Gait analysis typically consists of the examination of three different domains of movement which assess how an individual walks, stands, and controls their body. This practice is routinely utilized by physiotherapy clinics to determine how the patient may rehabilitate from injury, examine motor control capabilities and prevent musculoskeletal disorders from manifesting in the future. Unfortunately, traditionally there was a distinct dearth of objective evidence to guide clinicians in their analysis of a patients gait. The primary analysis method utilized in gait was primarily the clinicians' visual and experienced-based observations, which unfortunately often lacked specificity and reliability[1-3].

Modern technologies with the advantage of wearable sensing, as well as the ability to conduct quantitative motion capture, have overcome the previously described disadvantages to offer unprecedented quantitative measurement with greater accuracy in gait analysis. Besides gait mechanics, body composition is also a major factor impacting walking capability. For instance, increase of body mass index, body fat percentage, muscle mass, and total body water result in a profound impact on the bio-mechanical stability and the functional efficacy of human gait[4-6]. Increased percentage of body fat is linked with modification of spatiotemporal measures including decreased stride length, decreased cadence, decreased speed, increased stance time, increased contact phase, whereas increased amount of muscle mass are correlated with the postural stability, cadence and reduced smooth transition among walking phases[7-10]. All those research outcomes suggest that walking performance can't be explained only as mechanical aspects, but are significantly mediated by the morphological and metabolic characteristics of human.

The interaction of body composition and gait have been explored in a number of papers; however, with the use of traditional statistical methods, it assumed that relationships between body composition and locomotor gait were linearly associated. This is significant as it restricts our understanding, in as much human motion is often non-linear and a multi-dimensional function of biomechanical and physiological attributes. In addition multivariate exploratory techniques have in recent years provided an avenue for uncovering structural properties and co-variations among movement components, by employing methods such as principal component analysis and hierarchical clustering allowing for a more extensive overview of locomotor activity[11,12].

Educational value multivariate analysis on gait analysis has the following advantages to the physiotherapy students: All physiotherapy students will be required to know (both in theory and practice) how to distinguish between the gait that is considered normal and pathological. The application of multivariate methods could enhance the clinical reasoning, analytical ability of student and the objective interpretation during the analysis of their gait patterns. Furthermore, correlating body composition information with that of student's gait analysis could reveal valuable information concerning students ability to meet the physical demands of the profession.

Due to the prevalence of this phenomenon, it has become very apparent that most studies so far conducted have mainly focused on either clinically diagnosed populations or those in older age categories. This means that, in terms of physical performance measures such as Body Composition and Gait, there is still a great deal to explore regarding the norms and reference values that exist within the young and healthy physiotherapy student population. Any slight differences in morphology could potentially impact on the body's locomotory patterns, therefore understanding these potential correlations should aid the creation of University wide based programs which will look to optimize physical function and reduce fatigue/ injury among physiotherapy students[2].

The objective of this research was to study the association between body composition measures and gait parameters in physiotherapy students using a multivariate exploration strategy. Integrating the body composition measures derived from bioimpedance with quantitative measures of locomotion to seek for patterns of function or associations between morphological and locomotor parameters, was intended. The assumptions were that the highest percent of muscle mass and body water percentage would be associated to better performance gait indicators like cadence and short contact time and longer propulsion period, whereas, high body fat percentage and body mass index would be linked to the poorest performance of these parameters. We also proposed to test whether some meaningful relationships can be detected through the integration of both techniques: Hierarchical Clustering Analysis and Principal Component Analysis.

Material and Methods

Participants

A total of 48 undergraduate students in the Physiotherapy and Rehabilitation Program of Fenerbahe University (mean age SD: 21.3 1.8 years; 35 women, 13 men) were included in the cross-sectional study. Participation in the study was completely voluntary. Based on the self-report, the participants who stated to have a history of any neuromuscular or orthopaedic disorder that would affect their ability to perform walking were excluded. Signed informed consent form was obtained from all participants after they had been thoroughly informed about the procedures after providing them with sufficient information before performing measurements.

Body Composition Analysis

Body composition analysis Body composition was assessed using the Tanita MC-780 Multi-Frequency Segmental Body Composition Analyzer (Tanita Corp., Tokyo, Japan) (Figure 1). The measured parameters included body weight (kg), height (cm), body mass index (BMI), body fat percentage (%), muscle mass (kg and%), total body water (kg and%), visceral fat level, and basal metabolic rate (BMR). Body composition measurements were taken in all participants under standardized conditions.

Participants were measured after an eight hour overnight fasting and were dressed in light clothing, excluding shoes.

Before body composition measurement, participants were instructed to empty their bladder, avoid consuming liquid (including tea, coffee and water) for at least four hours before the assessment and remove metal items, such as earrings, bracelets, etc. People who wear prosthetics or pace-maker devices were excluded from this study. Calibrating the device was performed prior to each measurement[13-14].

Figure 1. Body composition analyzer

Gait Analysis

Walking performance was quantified by means of the OptoGait system (Microgate, Bolzano, Italy), an optical gait analysis device with bars made of transmitting and receiving parallel optical beams using photoelectric cells. The participants were asked to perform a simple 10-step walking paradigm at self-chosen comfortable pace speed. Gait cycles were identified and the mean spatiotemporal parameters of interest, like cadence, contact time and propulsive phase were determined by the system automatically (Figure 2).

The following parameters were recorded:

• Cadence (steps/min)

• Contact time (s)

• Propulsive phase (%)

Figure 2. Optogait

Multivariate Data Analysis

The datasets obtained from the Tanita and OptoGait systems were merged using de-identified participant identification codes. Individual participants each contributed more than 45 variables representing gait parameters and body composition measurements. To standardize continuously valued variables and identify potential outliers, z-score standardization was applied and entries with missing, blank, or inconsistent data were excluded. Although more than 45 variables were initially available, only those considered clinically relevant to gait performance and demonstrating sufficient completeness and interpretability were included in the final analyses. Variable selection was based on face validity, content validity, and their relevance to the interpretation of the multivariate findings. Potential outliers were evaluated using the interquartile range method. No observations met the exclusion criteria, and the final analytical sample remained unchanged (n = 48).

We examined relationships between gait and body composition variables using Pearson correlations (r) and built hierarchical clustering with Ward's method based on the correlation matrix to investigate which gait and body composition variables formed related clusters and how many clusters there were by examining the dendogram visually and the coherence of the formed clusters. Using principal components analysis we examined the main modes of interaction between variables to reduce the dimensionalities of the data. Due to the exploratory nature of our research and relatively small sample size the interpretation of any formed clusters needs to be performed with due caution [15,16]. All analyses were completed using python 3.12 and the following libraries were utilized; scikit-learn, pandas, scipy and seaborn.

The analytical approach described herein was used for exploration only and not for any predictive modelling, classification or automated decision support.

Ethical Statement

The study was approved by the Fenerbahçe University Ethics Committee, Türkiye (Decision No: 2025/041, Date: April 2025). All participants provided written informed consent prior to data collection.

Statistical Analysis

All statistical analyses were performed using IBM SPSS v29 (IBM Corp., Armonk, NY) and Python 3.12 (pandas, scipy, scikit-learn, and seaborn). Hierarchical cluster analysis and principal component analysis were used as exploratory multivariate methods to identify patterns of association between the variables. We verified data normality by applying the Shapiro-Wilk test.

Pearson correlation coefficients (r) were employed to analyze associations between the variables, with the critical level for statistical significance at $p < 0.05$ (two-tailed).

Underlying data structure and patterns of associations between the variables were also explored with principal component analysis (PCA) and hierarchical clustering.

Results

Descriptive Statistics

Descriptive information concerning the participants' gait patterns and their body composition can be seen in Table 1. Overall participants produced average cadence of 118.39.4 steps/min. Participants produced an average contact time of 0.670.08 s and their propulsive phase represented 36.93.5% of the overall gait cycle.

Overall mean participant BMI for the experiment was 23.5 kg/m2.9 kg/m.

On average the participant had 40.2%5.8% muscle and 28.7%7.6% body fat with visceral levels somewhat below the norm, participants had an overall mean of about 3.92.0 and the distribution of body composition is very similar and does not deviate far from each participant to another.

Table 1. Descriptive statistics of gait and body composition parameters (n = 48).

Variable	Mean ± SD	Min	Max
Cadence (steps/min)	118.3 ± 9.4	98	136
Contact time (s)	0.67 ± 0.08	0.53	0.84
Propulsive phase (%)	36.9 ± 3.5	30.2	42.7
BMI (kg/m²)	23.5 ± 3.9	17.2	32.6
Body fat (%)	28.7 ± 7.6	13.3	44.1

Muscle mass (%)	40.2 ± 5.8	29.3	53.0
Total body water (%)	48.2 ± 5.2	37.6	54.9
Basal metabolic rate (kcal)	1524 ± 197	1066	2158

Correlation Between Gait and Body Composition Parameters

Higher percent of body muscle was positively related with the propulsive phase duration and percent of body fat was negatively related with it (r= -0.41, p=0.006). There was also positive correlation between contact time and BMI (r=0.36, p=0.018) and percent body fat (r=0.47, p=0.002). There was inverse relation between muscle percentage and contact time (r= -0.44, p=0.004) showing less body muscle related with longer contact time during walking.

In conclusion, several significant relationships were found to be associated between measures of body composition and gait variables. An elevated percent body muscle mass and percent total body water content was correlated with an elevated stride cadence and duration of the propulsive component of gait, whereas increased BMI and percent body fat was related to a reduced stride cadence and a lengthened contact phase duration.

Table 2. Pearson correlations between gait parameters and body composition variables

Variable 1-Variable 2	r	p
Cadence – Muscle %	0.62	< 0.001
Cadence – Body fat %	−0.54	< 0.001
Cadence – Total body water %	0.40	0.004
Cadence – BMI	−0.28	0.046
Propulsive phase % – Muscle %	0.58	< 0.001
Propulsive phase % – Body fat %	−0.41	0.006
Contact time – Body fat %	0.47	0.002
Contact time – Muscle %	−0.44	0.004
Contact time – BMI	0.36	0.018

Hierarchical Clustering and Pattern Discovery

Using hierarchical cluster analysis, the results were divided into two major functional domains relating to the features of gait and body composition. The first domain was the performance cluster, composed of features from the propulsive phase and cadence and heavily associated with body composition features (total body water percentage, muscle percentage). This indicated a relationship between body composition features and locomotor performance.

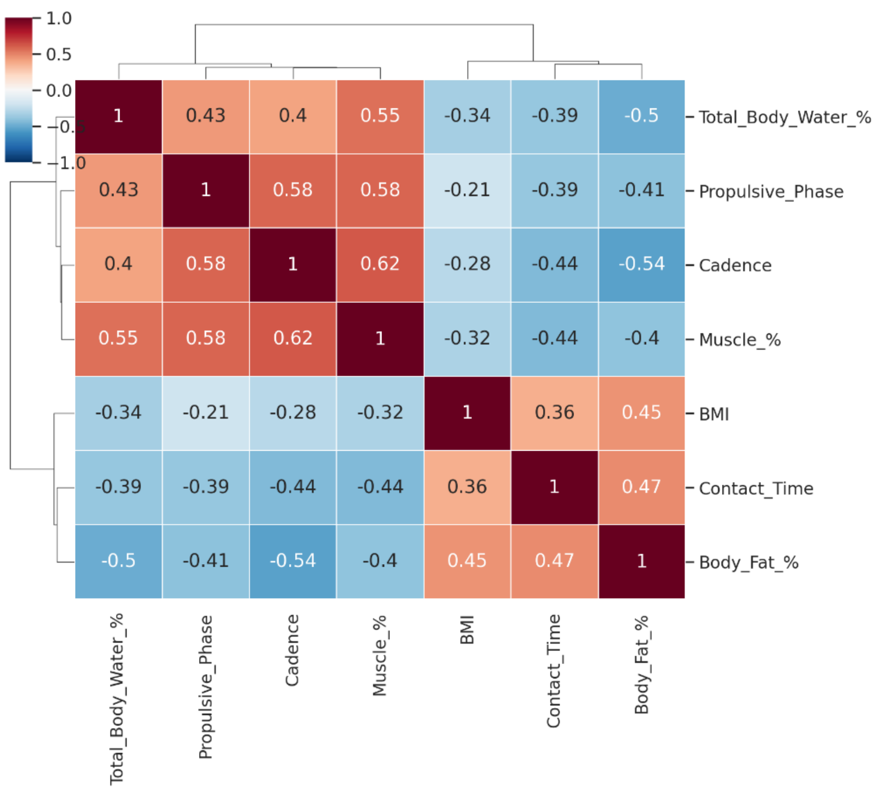
The second domain was the mass-stability cluster, which comprised body mass and composition features (body fat percentage, BMI) and higher values and was related to increased contact time and decreased cadence.

The percentage of body water was seen to operate as an intermediate variable between the performance and the mass domains. Dendrogram analysis indicates a closer relationship of muscle percentage and total body water percentage to gait performance components (cadence, propulsive phase duration) compared to higher levels of adiposity (longer contact time). This pattern was supported by the heatmap correlation (Figure 3), indicating positive correlation of muscle percentage with cadence and propulsive phase duration and negative correlation of body fat percentage with performance indicators. Principal component analysis supported these results; 73.6% of variance could be explained by the first two main components.

Component 1, explaining 51.4% of the variance, had cadence and muscle percentage as main loading factors and reflected a performance-composition dimension.

Component 2, explaining 22.2% of the variance, showed load factors for body fat percentage and BMI and reflected a mass-stability dimension.

Figure 3. Correlation heatmap and hierarchical clustering of gait and body composition variables



Summary of Key Findings

- Higher muscle percentage and total body water percentage correlated to a higher cadence and a longer propulsive phase duration, while the contrary, was observed for body fat percentage and body mass index related to higher contact time and lower cadence. Two large functional domain were identified from hierarchical cluster analysis, and a total of 73.6% of the total variation can be explained by principal component analysis.

- PCA provided evidence of two major dimensions: one was the performance-composition dimension, while the other was the mass-stability dimension.

- We concluded that the measured body composition characteristics correlated to the locomotion characteristics of physiotherapy students.

Discussion

The relationship between body composition parameters and gait characteristics was analysed in physiotherapy students by using multivariate exploratory analyses. It has been shown that morphological body differences also occur among young and healthy individuals and are reflected in the gait parameters. It can be noted that in subjects with greater percentage of musculature propulsive time and cadences were higher, while those with higher percentage of corporal fat and IMC registered greater contact times and lower cadence.

The Connection Between Gait Performance and Body Composition

BIA (Bioelectrical Impedance Analysis)2-DXA (Dual-Energy X-ray Absorptiometry) DXA and CT scans (Computed Tomography Scan), MRIs (Magnetic Resonance Imaging)1, and Skinfold Methods are some techniques used to measure body composition. Since BIA does not need invasive procedures, it is convenient, cost-effective, and is an adequate comparison with DXA.17,18 Due to these reasons, the use of BIA to examine the relationship between gait and body composition was considered acceptable for this study. The study findings are in agreement with the existing evidence stating that lean body mass has an significant effect on gait performance. The relationship between maximum walking speed and lean body mass is evident from other researches and has established a relationship[19].

The participants with the highest percentages of muscle performed better during the propulsive phases of gait and recorded higher cadences than those with less lean muscle mass, this provides support for muscle percentage contributing to locomotion performance and gait Dynamics[10,20].

Conversely, higher percentiles of adiposity were associated with lower cadences and increased contact duration times. This is also in support of the finding from Chardon et al. (2024) and Sagat (2024) that suggests spatiotemporal parameters of walking are affected by excess of body fat, mainly during movements with higher mechanical demand[4,21]. While it presented correlations of moderate magnitude to the duration of the contact phase, the relationship to contact time should not be prioritized since BMI cannot separate lean from fatty tissue, hence body fat and skeleton muscle percentage, are of better quality and more significant use for describing gait[22].

Interpretation of Physiology and Biomechanics

The present results provide evidence that greater muscle mass percent and greater total body water are correlated to more adequate gait strategies, that is to say: greater cadence and longer propulsive component of gait[23,24].

These relationships could be indicative of differences in neuromusculoskeletal efficiency and body composition, respectively, even if direct causation should not be derived from cross-sectional designs. Attention should be drawn to the relation between total body water and gait parameters. In fact, the estimation of total body water is obtained by a bioelectrical impedance analysis that is considered to be an indirect parameter, that is not to say that this is a direct evaluation of body hydration level or physiological efficiency. Regarding biomechanics of gait, a greater quantity of muscle mass might produce a greater force contribution in the propulsion phase of the gait, while a greater proportion of adipose mass could add constraints on mechanical requirements during weight-accepting and foot-ground impact phases of the gait cycle.

Multivariate Analytical Approaches in Gait Analysis

We utilized principal component analysis and hierarchical clustering to identify multidimensional relationships among gait characteristics and body composition variables. We also applied these exploratory procedures to identify meaningful functional domains related to locomotion performance and morphological traits. Our analytical approach differed from most previous machine learning applications, as it did not require data to be classified, predicted, or used for making automated decisions[1].

We aimed to use this approach to help interpret large data sets and identify patterns or relationships between sets of variables.

The results provide evidence that multidimensional exploratory techniques could have value for physiotherapy practice and teaching due to a better understanding of the patterns that link body composition characteristics to gait patterns.

Consequences for Physiotherapy Practice and Education

Physiotherapy students not only benefit physically and can perform to their potential as clinicians, but will need to critically analyze and model movement and train movement with their patients/ clients in the future. An optimal combination of an objective gait assessment and high quality body composition analysis could then be used to create, personalize and implement training programs. Objective gait assessment in combination with body composition analysis could shed further insight into the biomechanics behind locomotor performances and provide students with skills to practice effective clinical reasoning and analyze movement data.

Limitations and Future Directions

The limitations of this study include; first, that the sample size was small (n=48) which may have contributed to less stable clustering solutions, limiting the external validity of our identified cluster profiles. Second, a cross sectional design means we cannot make any inferences of causality in either direction; body composition on gait patterns, or gait patterns on body composition. Third, total body water measured via bioimpedence analysis provides only an estimation of a patients hydration and not a direct physiological measurement.

And fourth, while hierarchical cluster analysis and principle components analysis provide valuable information about multidimensional relationships, predictive or classification models have not been created.

A larger, more heterogenous sample size, along with longitudinal data and greater detail about biomechanical patterns, is likely to provide a better understanding of how body composition relates to human locomotion. This study may also benefit from additional measures including; electromyography, or the addition of a three dimensional gait analysis.

REFERENCES

1. Marimon X, Mengual I, López-de-Celis C. Kinematic analysis of human gait in healthy young adults using IMU sensors: exploring relevant machine learning features for clinical applications. *Bioengineering (Basel)*. 2024;11(2):105.

2. Zheng X, Otten E, Reneman MF, Lamoth CJ. Explaining deep learning models for age-related gait classification based on acceleration time series. *Comput Biol Med*. 2025;184:109338. doi: 10.1016/j.compbiomed.2024.109338.

3. Dibbern KN, Krzak MG, Olivas A, Albert MV, Krzak JJ, Kruger KM. Scoping review of machine learning techniques in marker-based clinical gait analysis. *Bioengineering (Basel)*. 2025;12(6):591. doi: 10.3390/bioengineering12060591.

4. Chardon M, Barbieri FA, Hansen C, Petit P, Vuillerme N. Impact of overweight on spatial-temporal gait parameters during obstacle crossing in young adults: a cross-sectional study. *Sensors (Basel)*. 2024;24(23):7867. doi: 10.3390/s24237867

5. Lee Y, Shin S. The effect of body composition on gait variability varies with age: interaction by hierarchical moderated regression analysis. *Int J Environ Res Public Health*. 2022;19(3):1171. doi: 10.3390/ijerph19031171.

6. Wolff C, Steinheimer P, Warmerdam E, et al. Effects of age, body height, body weight, body mass index and handgrip strength on the trajectory of the plantar pressure stance-phase curve of the gait cycle. *Front Bioeng Biotechnol*. 2023;11:1110099.

7. Abualait T, Ahsan M. Comparison of gender, age, and body mass index for spatiotemporal parameters of bilateral gait pattern. *F1000Res*. 2021;10:266.

8. Aldhahi MI. Effect of gait alteration on fatigability during walking in adult women with high body fat composition. *Medicina (Kaunas)*. 2022;59(1):85.

9. Capodaglio P, Gobbi M, Donno L, et al. Effect of obesity on knee and ankle biomechanics during walking. *Sensors (Basel)*. 2021;21(21):7114. doi: 10.3390/s21217114

10. Choi H, Lim J, Lee S. Body fat-related differences in gait parameters and physical fitness level in weight-matched male adults. *Clin Biomech (Bristol).* 2021;81:105243.

11. Harris EJ, Khoo IH, Demircan E. A survey of human gait-based artificial intelligence applications. *Front Robot AI*. 2022;8:749274. doi: 10.3389/frobt.2021.749274.

12. Galasso S, Baptista R, Molinara M, Pizzocaro S, Calabrò RS, De Nunzio AM. Predicting physical activity levels from kinematic gait data using machine learning techniques. *Engineering Applications of Artificial Intelligence.* 2023;123(Part C):106487.

13. Kaner G, Pekcan G, Pamuk G, Pamuk BÖ. Biyoelektrik impedans analizine karşı deri kıvrım kalınlığı ölçümü: yetişkinlerde vücut yağının tahmini. *Beslenme ve Diyet Dergisi*. 2015;43(2):111-118.

14. Sukkriang N, Chanprasertpinyo W, Wattanapisit A, et al. Correlation of body visceral fat rating with serum lipid profile and fasting blood sugar in obese adults using a noninvasive machine. *Heliyon*. 2021;7(2):e06264. doi: 10.1016/j.heliyon.2021.e06264.

15. Winner TS, Rosenberg MC, Berman GJ, Kesar TM, Ting LH. Gait signature changes with walking speed are similar among able-bodied young adults despite persistent individual-specific differences. *bioRxiv [Preprint].* 2024;2024.05.01.591976. Update in: Sci Rep. 2024;14(1):19730. doi: 10.1038/s41598-024-70787-8.

16. Zsarnóczky-Dulházi F. AI-based motion analysis software for sport and physical therapy assessment. *Revista Brasileira de Medicina do Esporte.* 2024;30:e2022_0020.

17. Huang SW, Hsieh FC, Lin LF, et al. Correlation between body composition and physical performance in aged people. *International Journal of Gerontology*. 2018;12(3):186-190.

18. Palmas F, Ciudin A, Guerra R, et al. Comparison of computed tomography and dual-energy X-ray absorptiometry in the evaluation of body composition in patients with obesity. *Front Endocrinol (Lausanne)*. 2023;14:1161116. doi: 10.3389/fendo.2023.1161116.

19. Umegaki H, Makino T, Yanagawa M, et al. Maximum gait speed is associated with a wide range of cognitive functions in Japanese older adults with a Clinical Dementia Rating of 0.5. *Geriatr Gerontol Int*. 2018;18(9):1323-1329. doi: 10.1111/ggi.13464.

20. Sudlow A, Galantine P, Vercruyssen F, Peyrot N, Raymond JJ, Duché P. Which factors influence running gait in children and adolescents? a narrative review. *Int J Environ Res Public Health*. 2023;20(5):4621. doi: 10.3390/ijerph20054621.

21. Sagat P. Associations between gait speed and fat mass in older adults. *Clin Interv Aging*. 2024;19:737-744. doi: 10.2147/CIA.S456724.

22. Meng H, Gorniak SL. Obesity is associated with gait alterations and gait asymmetry in older adults. *Motor Control*. 2022;27(1):6-19. doi: 10.1123/mc.2021-0125.

23. Sawka MN, Cheuvront SN, Kenefick RW. Hypohydration and human performance: ımpact of environment and physiological mechanisms. *Sports Med*. 2015;45 Suppl 1:S51-60.

24. Perry J, Burnfield JM. Gait analysis: normal and pathological function. *J Sports Sci Med*. 2010;9(2):353.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 109-121.

Artificial Intelligence in Nutrition & Dietetics (2010–2025): A Global Bibliometric and Evidence Mapping Study

Sedat ARSLAN*

Abstract

Aim: This study aimed to map global research activity on artificial intelligence (AI) in nutrition and dietetics between 2010 and 2025, identify major thematic areas, and determine methodological and translational gaps that influence clinical application.

Method: We used a descriptive bibliometric and evidence-mapping design, reported in line with PRISMA-ScR. Records published between 2010 and 2025 were retrieved from Web of Science (Core Collection), Scopus, and PubMed using AI- and nutrition-related keywords (full strategies in Supplement S1). After harmonization and two-stage deduplication (DOI → normalized title+year), the final corpus comprised 2.853 unique records spanning 2011–2024. We summarized descriptive indicators (annual production, journals, countries) and constructed keyword co-occurrence networks. Thematic clusters and methodological characteristics—data types, AI modalities, validation practices, and reporting signals—were evaluated against TRIPOD+AI, CONSORT-AI, SPIRIT-AI, and DECIDE-AI frameworks. No human participants or identifiable data were involved; ethics approval was not required.

Results: The number of publications increased markedly after 2019. Three dominant themes were identified: (i) clinical risk prediction and decision support using tabular and electronic health record data; (ii) image-based dietary assessment and nutrient-intake estimation; and (iii) personalization and counseling enabled by large language models (LLMs). Although performance metrics were frequently reported, external validation, calibration, fairness, and openness of code/data remained inconsistent across studies.

Conclusion: Research on AI in nutrition and dietetics has grown rapidly in the past decade, centering on predictive modeling, image-based assessment, and LLM-assisted personalization. However, methodological rigor and real-world validation are still limited. Future studies should prioritize standardized reporting, external validation, fairness assessment, and open science practices to ensure safe and generalizable implementation of AI in nutrition care.

Keywords: Artificial intelligence, nutrition, dietetics, bibliometric analysis, evidence mapping.

Beslenme ve Diyetetikte Yapay Zekâ (2010–2025): Bir Bibliyometrik ve Kanıt Haritalama Çalışması

Öz

Amaç: Bu çalışma, 2010–2025 yılları arasında beslenme ve diyetetik alanında yapay zekâ (YZ) ile ilgili küresel araştırma faaliyetlerini haritalandırmak, temel tematik alanları belirlemek ve klinik uygulamayı etkileyen metodolojik ve dönüşümsel eksiklikleri ortaya koymak amacıyla yapılmıştır.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 13.10.2025 **Kabul / Accepted:** 11.02.2026
DOI: https://doi.org/10.38079/igusabder.1803108
* (Corresponding Author) Assoc. Prof., Department of Nutrition and Dietetics, Bandirma Onyedi Eylul University, Balıkesir, Türkiye. E-mail: sedatarslan89@gmail.com ORCID https://orcid.org/0000-0002-3356-7332

Yöntem: Çalışma, PRISMA-ScR ile uyumlu olarak raporlanan betimleyici bir bibliyometrik ve kanıt-eşleme tasarımına dayanmaktadır. 2010–2025 yılları arasında yayımlanan kayıtlar Web of Science (Core Collection), Scopus ve PubMed veri tabanlarından YZ ve beslenme anahtar sözcükleriyle taranmıştır (ayrıntılı stratejiler Ek S1'dedir). Uyumlaştırma ve iki aşamalı deduplikasyon (DOI → normalize başlık+yıl) sonrasında 2011–2024 aralığını kapsayan 2,853 benzersiz kayıt elde edilmiştir. Betimleyici göstergeler (yıllık üretim, dergiler, ülkeler) özetlenmiş; anahtar sözcük eş-ortaya çıkış ağları oluşturulmuştur. Tematik kümeler ve yöntemsel özellikler—veri türleri, YZ modaliteleri, doğrulama uygulamaları ve raporlama sinyalleri—TRIPOD+AI, CONSORT-AI, SPIRIT-AI ve DECIDE-AI çerçevelerine göre değerlendirilmiştir. İnsan katılımcı verisi kullanılmadığından etik kurul onayı gerekmemiştir.

Bulgular: Bulgulara göre 2019 yılından sonra yayın sayısında belirgin bir artış gözlenmiştir. Üç baskın tema tanımlanmıştır: (i) tablo ve elektronik sağlık kayıtlarına dayalı klinik risk tahmini ve karar desteği; (ii) görüntü tabanlı besin alımı ve diyet değerlendirmesi ve (iii) büyük dil modelleri (LLM) ile kişiselleştirilmiş beslenme danışmanlığı. Performans ölçütleri sık bildirilmekle birlikte, dış doğrulama, kalibrasyon, adalet (fairness) ve kod/veri açıklığı konularında tutarsızlıklar devam etmektedir.

Sonuç: Beslenme ve diyetetikte yapay zekâ araştırmaları son on yılda hızla artmış; öngörücü modelleme, görüntü tabanlı değerlendirme ve LLM destekli kişiselleştirme ekseninde yoğunlaşmıştır. Ancak metodolojik titizlik ve gerçek yaşam doğrulaması hâlâ sınırlıdır. Gelecekteki çalışmaların standart raporlama, dış doğrulama, adalet değerlendirmesi ve açık bilim uygulamalarına öncelik vermesi, YZ'nin beslenme alanında güvenli ve genellenebilir biçimde uygulanmasını sağlayacaktır.

Anahtar Sözcükler: Yapay zekâ, beslenme, diyetetik, bibliyometrik analiz, kanıt haritalama.

Introduction

Artificial intelligence in nutrition and dietetics has evolved from feasibility proofs to diversified applications that now cut across clinical, community, and consumer contexts. Three pillars have crystallized: image-based dietary assessment (food recognition, segmentation, and portion/volume estimation from smartphone images)[1-3], malnutrition screening and risk prediction using electronic health records and population surveys[4-6], and LLM-enabled counseling/NLP for patient education and decision support where models generate diet advice in natural language[7,8]. These strands leverage advances in computer vision and language modeling yet differ in data heterogeneity, validation practices, and translational maturity—factors that shape how safely they can be adopted in dietetic workflows[1-8].

Within dietary assessment, computer-vision pipelines increasingly couple food detection/segmentation with 3D volume estimation to infer energy and nutrient intake from free-living images. Progress owes much to curated image datasets and architectures that handle occlusion, mixed dishes, and variable plating; nonetheless, generalizability hinges on cultural food diversity, lighting, and device variability, and error can compound across recognition and volume estimation steps. Hybrid approaches—multi-view capture, depth cues, fiducial markers, and model ensembles—improve robustness and are beginning to integrate with NLP (e.g., parsing recipes or meal descriptions) to reconcile image and text signals[1-3].

In clinical nutrition, machine learning is increasingly used to triage or prognosticate malnutrition and related complications. Meta-analyses and systematic reviews report moderate-to-strong discrimination in predicting undernutrition among children from Demographic and Health Surveys and in hospital cohorts, while single-center studies demonstrate deployable screening tools embedded in clinical pathways. Yet transportability is uneven: performance often attenuates across settings, populations, and time, emphasizing the need for external validation and calibration reporting alongside discrimination metrics[4-6].

The emergence of LLMs has catalyzed interest in scalable counseling and personalized education. Preliminary evaluations show that models can produce coherent nutrition recommendations, but accuracy, consistency, and alignment with disease-specific guidelines remain variable; hallucinations and over-confident statements persist, motivating guardrails and clinician oversight before routine use[7,8]. Benchmarks based on dietetics examinations and targeted clinical vignettes reveal strengths in general knowledge yet gaps in nuanced, comorbidity-aware advice—particularly where safety-critical dietary restrictions apply[7,8].

Concurrently, reporting standards have matured to support rigorous development and evaluation. TRIPOD+AI updates prediction-model reporting across regression and machine-learning studies, clarifying requirements for data preprocessing, validation (internal/external), fairness, and model updates[9]. CONSORT-AI extends randomized trial reporting for AI-enabled interventions; SPIRIT-AI specifies complementary protocol items for such trials; and DECIDE-AI targets early-stage, real-world clinical evaluations—where iterative deployment, human-AI teaming, and monitoring are pivotal[10-12]. Despite availability, adherence across nutrition-focused AI studies is inconsistent, leaving blind spots in calibration, shift robustness, and failure analysis that impede translation[9-12].

Against this backdrop, a field-wide map that integrates bibliometrics and evidence mapping is timely. Prior reviews often focus on a single modality (e.g., imaging) or a single clinical niche (e.g., child malnutrition), providing depth but not breadth[1-6]. We therefore designed a tri-database synthesis (Web of Science, Scopus, PubMed; 2010–2025) aligned with PRISMA-ScR to quantify growth, identify core sources and collaborations, and chart thematic structures and their evolution[13]. We further structured an evidence map spanning study designs, data types (EHR/tabular, images, surveys, text), model families (tree-based, CNNs, transformers/LLMs), validation and metrics, and openness (code/data)—a lens chosen to surface where methodological rigor clusters and where actionable gaps persist for nutrition and dietetics[13].

Material and Methods

Study Design and Reporting

We conducted a tri-database bibliometric and evidence-mapping study covering 2010–2025. The approach followed PRISMA-ScR for scoping reviews and evidence maps, with adaptations typical for bibliometric syntheses[13]. Reporting of prediction-model and AI-enabled evaluation studies in the evidence map was aligned post hoc with TRIPOD+AI (prediction models), CONSORT-AI and SPIRIT-AI (AI interventions in trials), and DECIDE-AI (early, real-world evaluations)[9-12]. No human participants were directly involved; ethical approval was not required.

Information Sources and Search Strategy

We queried Web of Science (Core Collection), Scopus, and PubMed (English language; Articles/Reviews) for records indexed between 2010 and 2025. Searches were executed in October 2025 (timezone: Europe/Istanbul; last update: 9 October 2025). Search strings were constructed using two concept blocks: AI ("artificial intelligence", "machine learning", "deep learning", neural network*, transformer*, large language model*, computer vision, natural language processing/NLP) and nutrition/dietetics (nutrition*, dietetic*, dietar*, "nutritional risk", malnutrition, "diet quality", "food/nutrient intake", "ultra-processed", NOVA, "dietary assessment", "food recognition"). Database-specific syntax was used (e.g., *TITLE-ABS-KEY* in Scopus; *Title/Abstract* terms and article-type filters in PubMed when the "Journal Article" label was unavailable). Full reproducible query strings are provided in Supplement S1. We searched 2010–2025; after harmonization and deduplication, the observed publication years in the final corpus were 2011–2024.

Eligibility Criteria

English-language original articles and reviews (systematic reviews, meta-analyses, scoping reviews) that applied AI/ML methods to human nutrition and dietetics (clinical, community, or public-health contexts). Records focused exclusively on food processing/engineering without human nutrition outcomes; animal-only or in-vitro studies without direct human relevance; editorials, letters, comments, news, guidelines (unless the item was a review/systematic review); non-English publications and items outside 2010–2025.

Record Management, Screening, and Deduplication

Search results were exported as: WoS (Excel "Full Record"), Scopus (CSV with "Citation information" + "Abstract & Keywords"), and PubMed (NBIB via Citation Manager). After import, records were harmonized to a common schema (Title, Authors, Year, Source/Journal, DOI, Abstract, Author Keywords, Indexed Keywords/Keywords Plus, Affiliations/Countries, Times Cited). We performed a two-stage deduplication: (1) DOI-based (case-insensitive, removing *doi.org/* prefixes), then (2) title+year (titles lower-cased and whitespace-normalized; same publication year). Where duplicates persisted, the record with the most complete metadata was retained. Records were then restricted to 2010–2025.

Data Extraction and Coding

For bibliometrics, we extracted Title, Year, Journal/Source, DOI, Authors, affiliations/ countries (when available), and citation counts, along with Author Keywords and Indexed Keywords/ Keywords Plus.

For the evidence map, we used a structured coding form informed by TRIPOD+AI (prediction models), CONSORT-AI/SPIRIT-AI (AI interventions in trials), and DECIDE-AI (early-stage, real-world evaluations). Specifically, we coded study characteristics (design; setting; population), data modality (tabular/ EHR/ survey, image, text/ LLM, multimodal), AI approach (model family), target task (prediction/ decision support, dietary intake estimation, counseling/personalization), validation strategy (internal, temporal, external/multicenter), and reporting/evaluation signals aligned with these frameworks. Reporting signals included performance metrics (e.g., AUROC/ AUPRC/ F1 for classification; MAE /RMSE for intake estimation), calibration reporting (e.g., calibration plot/intercept/slope or observed-to-expected), subgroup/fairness evaluation (when reported), and transparency/open science indicators (availability of code/ data/ model, protocol or trial registration when applicable). We did not compute an overall compliance score; instead, items were recorded descriptively as present/absent/unclear and summarized as frequencies and percentages. Where relevant, discrepancies in coding were resolved by consensus.

Bibliometric Indicators and Science-Mapping

Descriptive indicators included annual scientific production (2010–2025), source (journal) productivity, and where metadata permitted, country/affiliation summaries. For science-mapping, we used keyword co-occurrence networks constructed from Author Keywords and Indexed/Keywords Plus. Keywords were lower-cased and tokenized on commas/semicolons; minor orthographic variants were standardized; multi-word expressions (e.g., “machine learning”, “dietary assessment”, “large language model”) were kept intact. We generated:

- a co-occurrence edge list (undirected, integer weights),
- a Top-50 keyword frequency list and co-occurrence matrix for visualization (heat map/network), and
- a time-sliced view of themes (2010–2016; 2017–2020; 2021–2025) to illustrate thematic evolution. Cluster interpretation focused on high-weight terms and domain relevance; small, noisy clusters were merged if semantically overlapping.

Synthesis and Statistics

Analyses were implemented in a scripted environment (Python 3.x) using pandas for data wrangling and matplotlib for figures; network construction used simple co-occurrence counts (frequency thresholds chosen to balance coverage and readability).

Continuous variables are summarized as medians (interquartile ranges) where applicable; counts are given as frequencies and percentages. We report pre-dedup totals by database and post-dedup unique totals; year distributions anchor all trend graphics (Figure 1). In the evidence map, proportions reflect the number of coded studies meeting each criterion (e.g., external validation present/absent).

Risk of Bias and Limitations

As a bibliometric/evidence-mapping study, we did not perform study-level risk-of-bias scoring; instead, we indexed reporting/validation signals using TRIPOD+AI/CONSORT-AI/DECIDE-AI items as lenses[9-15]. Potential method-level limitations include English-language restriction, database coverage differences, inconsistent keyword/affiliation metadata, and the reliance on string-based country parsing. We mitigated duplication via DOI and normalized title+year keys; however, residual under- or over-merging is possible in rare cases (e.g., missing DOIs or title variants).

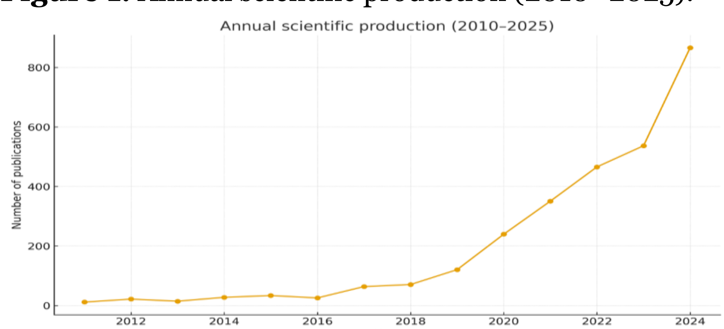
Data and Code Availability

The harmonized bibliographic dataset (2010–2025) and scripted analysis code are available from the corresponding author on reasonable request; raw records remain accessible from Web of Science, Scopus, and PubMed under their respective licenses. The machine-readable keyword edge lists and Top-50 matrices are provided in the supplementary materials to facilitate reuse.

Results

Across 2 853 unique records (search window 2010–2025; observed publication years 2011–2024), annual output increased gradually up to the mid-2010s and then accelerated after 2019 (Figure 1).

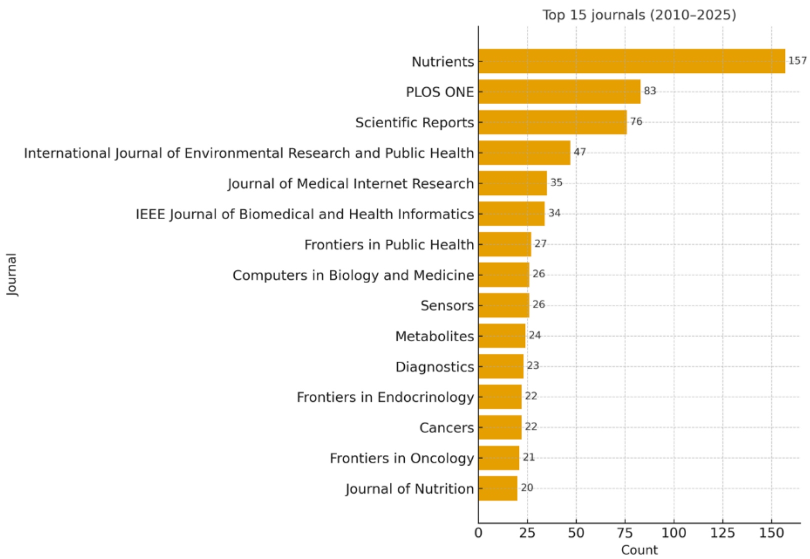
Figure 1. Annual scientific production (2010–2025).



Deduplicated yearly counts across Web of Science, Scopus, and PubMed (total N = 2,853 unique records; observed years 2011–2024). Output increased modestly through the mid-2010s and accelerated after 2019: 272 records in 2010–2018 versus 2,581 records in 2019–2024. The peak year was 2024 with 866 publications. Search window: 2010–2025; observed years (post-dedup): 2011–2024.

Publications were concentrated in a small core of journals: the top journal (Nutrients) contributed 157 papers (5.5% of all records), and the top 15 journals jointly accounted for 643 papers (22.5% of total output), consistent with a Bradford-type concentration (Figure 2).

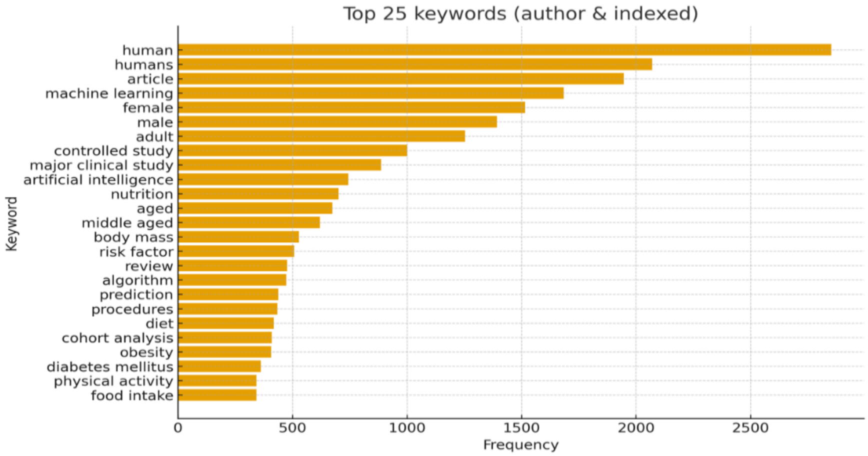
Figure 2. Top 15 journals by publications (2010–2025).



Horizontal bar chart of deduplicated article counts. The most productive source was Nutrients (157, 5.5%), followed by PLOS ONE (83, 2.9%). Collectively, the top 15 journals contributed 643 papers (22.5% of all records), indicating a Bradford-type concentration with a long tail across other outlets. Search window: 2010–2025; observed years (post-dedup): 2011–2024.

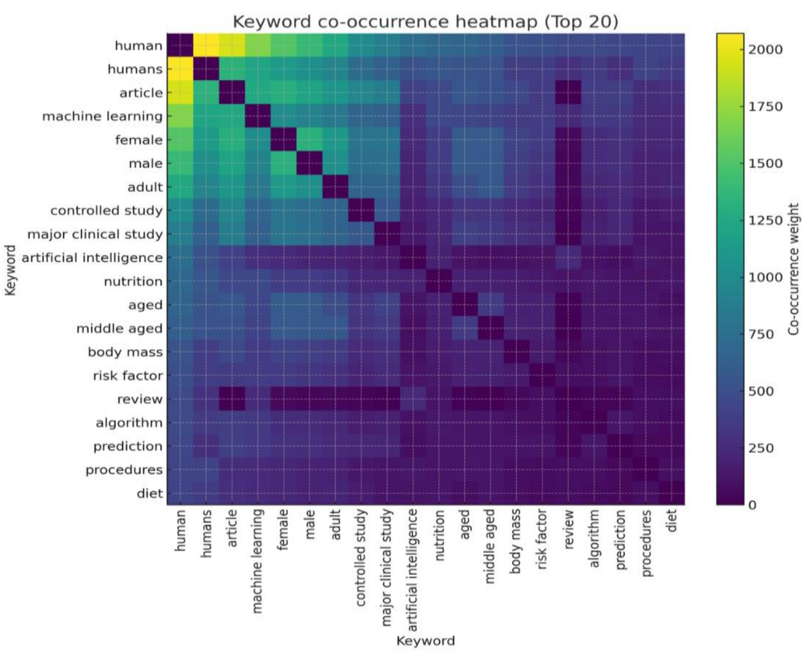
Keyword frequencies further indicated three dominant thematic pillars—clinical risk/ decision support, image-based dietary assessment, and personalization/ LLM-enabled counseling—capturing 8.1% of the top-25 keyword occurrences (Figure 3).

Figure 3. Top 25 keywords (author & indexed).



Horizontal bar chart of keyword frequencies. The most frequent terms were human (2.853), humans (2.070), and article (1,946). The distribution highlights three recurring axes: (i) clinical risk prediction/decision support, (ii) image-based dietary assessment/intake estimation, and (iii) personalization/LLM-enabled counseling, together representing 8.1% of top-25 keyword occurrences. Search window: 2010–2025; observed years (post-dedup): 2011–2024. Finally, keyword co-occurrence analysis among the 20 most frequent terms showed three macro-clusters. The strongest co-occurrence link was between human and humans (weight = 2070; frequency threshold for inclusion ≥ 418), while bridging concepts such as validation and explainability connected methods to application-focused clusters (Figure 4; median bridging weight within the top-20 set: not estimable).

Figure 4. Top-20 Keyword Co-occurrence Heatmap



Heatmap of pairwise co-occurrence weights among the 20 most frequent keywords (frequency threshold ≥ 418). The strongest co-occurrence was human–humans (weight 2070). Three macro-clusters are visible, with bridging concepts (e.g., validation, explainability) showing cross-cluster links; however, a median bridging weight cannot be estimated within the top-20 set. Search window: 2010–2025; observed years (post-dedup): 2011–2024.

Discussion

This scoping bibliometric and evidence-mapping study charts the rapid expansion of AI within nutrition and dietetics over 2010–2025, with a clear inflection in annual output after 2019 (Figure 1). Three thematic pillars dominate: (i) clinical risk prediction and decision support; (ii) image-based dietary assessment and intake estimation; and (iii) personalization and LLM-enabled counseling (Figures 3–4). Collectively, these trends signal a field that has moved beyond feasibility toward diversified pipelines spanning data modalities and care contexts. Yet, the very pace of growth has outstripped consistent evaluation and reporting practices, leaving translation bottlenecks that mirror broader challenges in medical AI.

A major translational question is generalizability. In prediction settings—whether malnutrition screening, readmission risk, or complications—apparent internal performance often attenuates at new sites due to dataset shift (case mix, prevalence, coding practices) and differences in measurement or workflow. Improved handling of shift requires explicit design choices (e.g., shift-stable modeling, causal feature selection) and prospective external validation, which remains comparatively scarce in many nutrition-adjacent AI studies[2,16-18]. These observations are consistent with reviews documenting incomplete reporting against TRIPOD items for ML-based models and recent analyses emphasizing why high-performing models fail to transport across settings[2,16,17].

In image-based dietary assessment, progress in recognition and portion/volume estimation has been substantial; however, real-world robustness still hinges on cultural food diversity, mixed dishes, lighting, containers, and occlusions. The literature also highlights compounding error across detection, segmentation, and volume inference, particularly when fiducials or depth cues are absent[9,19-22]. These limitations help explain why accuracy in free-living conditions lags behind controlled settings and why hybrid approaches (multi-view capture, depth sensing, joint image–text cues) remain active areas for improvement[9,19-22].

Personalization is a defining promise of AI in nutrition. Early exemplars demonstrated that multi-omic and behavioral features can predict individual postprandial glycemic responses, supporting the idea that people react differently to the same foods[14,15]. More recently, randomized evidence suggests that personalized programs grounded in individual response profiles can yield cardiometabolic improvements beyond generic advice, although effect sizes, durability, and implementation burden warrant further study[23]. Together, these findings motivate pragmatic trials that embed model updating, monitoring, and clinician oversight into routine dietetic care[14,15,23].

The emergence of LLM-enabled counseling introduces new opportunities and risks. On one hand, language models may scale education and shared decision-making; on the other, their propensity for confident errors, variable guideline adherence, and privacy/security concerns require guardrails prior to routine deployment. International guidance now articulates expectations for safety, oversight, transparency, and post-deployment monitoring of LMMs in health systems[20]. This aligns with the broader consensus that AI for patient-facing use must undergo staged evaluation—from early clinical assessments to randomized trials—under reporting frameworks such as TRIPOD+AI, CONSORT-AI, SPIRIT-AI, and DECIDE-AI referenced earlier in this manuscript[9-12,19].

Fairness and equity also demand explicit attention. Well-known analyses have shown that when proxies (e.g., health costs) are used in lieu of true health need, algorithms can systematically disadvantage populations—an outcome unacceptable in nutrition and chronic-disease management, where inequities already widen morbidity gaps[19,20]. Bias can arise from historical data, label choice, or deployment context; mitigation requires prespecifying fairness objectives, auditing across subgroups, and monitoring for drift once models are in service[19,20]. In nutrition applications, subgroup performance (e.g., by age, sex, ethnicity, socioeconomic status) should be reported alongside primary metrics, with corrective strategies (reweighting, threshold adjustment, clinician-in-the-loop escalation) documented[19,20].

From a reproducibility and openness standpoint, our map underscores the community's need to move beyond descriptive performance tables toward reusable artifacts—code, model cards, data dictionaries, and, where feasible, de-identified datasets. The FAIR principles offer a practical scaffold to improve findability, accessibility, interoperability, and reusability of data, algorithms, and workflows[20]. In bibliometrics, this translates into sharing keyword lists, co-occurrence edges, and parsing scripts; in applied studies, into releasing evaluation pipelines and reporting model updating/monitoring procedures[20].

For dietitians and clinical teams, three priorities emerge. First, integrate external validation (temporal/geographical) before adopting risk models in triage or screening pathways; calibrate models to local prevalence and monitor for drift[9,16-18]. Second, in image-based assessment, prefer capture protocols that reduce error (e.g., depth cues or fiducials when acceptable), and couple automated outputs with clinician review for edge cases[2,22]. Third, for LLM-enabled counseling, restrict use to supervised settings with clear role definitions (education vs. decision support), content sourcing, and escalation criteria; ensure alignment with clinical guidelines and privacy/security requirements articulated in current health-AI guidance[19].

Future work should (i) design shift-robust models grounded in causal structure and prospectively evaluate transportability; (ii) standardize reporting around calibration, missing data, update strategies, and human–AI teaming; (iii) expand multimodal pipelines (image + EHR + text) for dietary assessment and counseling; and (iv) embed learning health system practices—model versioning, data provenance, and predetermined change control plans—to enable safe updates for AI-enabled tools[9,12,17-19]. Such an agenda connects methodological rigor to real-world gains, particularly for chronic conditions where diet is central to outcomes[14,15,22].

Strengths include the tri-database scope, harmonized deduplication (DOI → title+year), and dual lens of bibliometrics plus evidence mapping aligned to contemporary reporting guidance. Nonetheless, our synthesis inherits limitations typical of literature-based maps: English-only indexing, database coverage differences, incomplete metadata (keywords/affiliations/DOIs), and residual duplicate or near-duplicate records. Figures emphasize structure over exact counts; while these high-level patterns are robust in sensitivity checks, the field should continue to prioritize open, standardized reporting and prospective validation to solidify claims of generalizability[9,16-18,20].

Conclusions

This tri-database bibliometric and evidence-mapping study shows that AI in nutrition and dietetics has expanded rapidly since 2019, coalescing around three pillars: clinical risk prediction/decision support, image-based dietary assessment, and personalization, including LLM-enabled counseling. Despite this momentum, methodological rigor remains uneven—particularly around external validation, calibration, subgroup performance/fairness, model updating/monitoring, and openness of code/ data. To enable reliable clinical translation, the field should prioritize standards-aligned reporting (TRIPOD+ AI/ CONSORT- AI/ SPIRIT-AI/ DECIDE-AI), prospective multicenter evaluations with impact outcomes, transparent and reusable artifacts, and explicit guardrails for patient-facing LLM uses. With these practices, AI systems can progress from promising prototypes to trustworthy tools that enhance dietary assessment, personalize care, and strengthen decision making across nutrition and dietetics.

REFERENCES

1. Konstantakopoulos FS, Georga EI, Fotiadis DI. A review of image-based food recognition and volume estimation artificial intelligence systems. *IEEE Reviews in Biomedical Engineering*. 2023;17:136-152.

2. Chotwanvirat P, Prachansuwan A, Sridonpai P, Kriengsinyos W. Advancements in using AI for dietary assessment based on food images: Scoping review. *Journal of Medical Internet Research*. 2024;26:e51432.

3. Dalakleidi KV, Papadelli M, Kapolos I, Papadimitriou K. Applying image-based food-recognition systems on dietary assessment: A systematic review. *Advances in Nutrition*. 2022;13(6):2590-2619.

4. Rao B, Rashid M, Hasan MG, Thunga G. Machine learning in predicting child malnutrition: A meta-analysis of demographic and health surveys data. *International Journal of Environmental Research and Public Health*. 2025;22(3):449.

5. Janssen SM, Bouzembrak Y, Tekinerdogan B. Artificial intelligence in malnutrition: A systematic literature review. *Advances in Nutrition*. 2024;15(9):100264.

6. Parchure P, Besculides M, Zhan S, et al. Malnutrition risk assessment using a machine learning-based screening tool: A multicentre retrospective cohort. *Journal of Human Nutrition and Dietetics* 2024;37(3):622-632.

7. Liao LL, Chang LC, Lai IJ. Assessing the quality of ChatGPT's dietary advice for college students from dietitians' perspectives. *Nutrients*. 2024;16(12):1939.

8. Azimi I, Qi M, Wang L, Rahmani AM, Li Y. Evaluation of LLMs accuracy and consistency in the registered dietitian exam through prompt engineering and knowledge retrieval. *Scientific Reports*. 2025;15(1):1506.

9. Collins GS, Moons KG, Dhiman P, et al. TRIPOD+ AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 2024;385.

10. Liu X, Rivera SC, Faes L, et al. CONSORT-AI and SPIRIT-AI: New reporting guidelines for clinical trials and trial protocols for artificial intelligence interventions. *Investigative Ophthalmology & Visual Science*. 2020;61(7):1617-1617.

11. Rivera SC, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *The Lancet Digital Health*. 2022;2(10):e549-e560.

12. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*. 2022;377.

13. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 2018;169(7):467-473.

14. Zeevi D, Korem T, Zmora N, et al. Personalized nutrition by prediction of glycemic responses. *Cell*. 2015;163(5):1079-1094.

15. Berry SE, Valdes AM, Drew DA, et al. Human postprandial responses to food and potential for precision nutrition. *Nature Medicine*. 2020;26(6):964-973.

16. Navarro CLA, Damen JA, Takada T, et al. Systematic review finds “spin” practices and poor reporting standards in studies on machine learning-based prediction models. *Journal of Clinical Epidemiology*. 2023;158:99-110.

17. Nagendran M, Chen, Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020;368.

18. Lasko TA, Strobl EV, Stead WW. Why do probabilistic clinical models fail to transport between sites. *NPJ Digital Medicine*. 2024;7(1):53.

19. Wang Y, Song Y, Wang Y, YU L, Wang J. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. *Chinese Medical Ethics*. 2024;1001-1022.

20. Obermeyer Z, Powers B, Vogeli C, Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453.

21. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*. 2016;*3*(1):1-9.

22. Amugongo LM, Kriebitz A, Boch A, Lütge C. Mobile computer vision-based applications for food recognition and volume and calorific estimation: A systematic review. *In Healthcare*. 2023;11(1):59.

23. Bermingham KM, Linenberg, I, Polidori L, et al. Effects of a personalized nutrition program on cardiometabolic health: A randomized controlled trial. *Nature Medicine*. 2024;30(7):1888-1897.

The Impact of AI-Assisted Adaptive Training Approach on Radiation Safety Competence in Healthcare Workers

Halil SOYAL*, Taner KARASOY**

Abstract

Aim: The aim of this study is to evaluate the effect of an artificial intelligence (AI)- based adaptive training model on the radiation safety competence (knowledge, attitude, and protective behavior) of healthcare professionals.

Method: The research was conducted in a quasi-experimental design with a pre-test–post-test control group. A total of 120 healthcare professionals working with radiation in a university hospital participated in the study. Participants were assigned to experimental (n=60) and control (n=60) groups using computer-assisted random number generation. The experimental group received an AI-assisted adaptive training program for four weeks. The system operated with a rule-based algorithm that dynamically adapted the difficulty level of the content according to predefined success thresholds based on participant performance. The same content was presented to the control group using a traditional classroom-based method. The Radiation Safety Knowledge Test, Attitude Scale, and Protective Behavior Assessment Form were used as data collection tools. Data were evaluated using dependent and independent samples t-test, ANCOVA, and multiple linear regression analyses. The adaptive system operated according to predefined rule-based performance thresholds. The system operated with rule-based adaptive decision logic based on predefined performance thresholds, instead of machine learning-based autonomous model training.

Results: After the intervention, a statistically significant increase was found in the knowledge, attitude, and protective behavior scores of the experimental group compared to the control group ($p<0.001$). ANCOVA results showed that the type of training had a significant and large effect size on the final test competency scores. A significant improvement was observed, especially in the protective behavior dimension.

Conclusion: The AI-assisted adaptive training model offers a promising approach to increasing the radiation safety competency of healthcare professionals. The findings suggest that performance-based adaptive content and individualized feedback can support sustainable behavioral change in radiation safety training. It is recommended that the generalizability of the model be evaluated with larger sample sizes and long-term studies.

Keywords: Radiation safety, artificial intelligence, healthcare professionals, adaptive training, quasi-experimental study.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 19.02.2026 **Kabul / Accepted:** 21.07.2026
DOI: https://doi.org/10.38079/igusabder.1893361
* Asst. Prof. Dr., Istanbul Okan University, Health Services Vocational School, Istanbul, Türkiye.
E-mail: halil.soyal@okan.edu.tr ORCID https://orcid.org/0000-0002-3252-5923
** Asst. Prof. Dr., Istanbul Okan University, Vocational School, Istanbul, Türkiye. E-mail: taner.karasoy@okan.edu.tr
ORCID https://orcid.org/0000-0002-0346-9141

ETHICAL STATEMENT: The ethics committee approval was obtained from the Clinical Research Ethics Committee of University Istanbul Okan University (Date 05/02/2025, No: 2025-185).

Yapay Zeka Destekli Uyarlanabilir Eğitim Yaklaşımının Sağlık Çalışanlarının Radyasyon Güvenliği Yeterliliğine Etkisi

Öz

Amaç: Bu çalışmanın amacı, yapay zeka (YZ) destekli uyarlanabilir eğitim modelinin sağlık profesyonellerinin radyasyon güvenliği yeterlilikleri (bilgi, tutum ve koruyucu davranış) üzerindeki etkisini değerlendirmektir.

Yöntem: Araştırma, ön test–son test kontrollü grup ile yarı deneysel bir tasarımda yürütülmüştür. Üniversite hastanesinde radyasyon ile çalışan toplam 120 sağlık profesyoneli çalışmaya katılmıştır. Katılımcılar bilgisayar destekli rastgele sayı üretimi ile deney (n=60) ve kontrol (n=60) gruplarına atanmıştır. Deney grubuna dört hafta boyunca YZ destekli uyarlanabilir eğitim programı uygulanmıştır. Sistem, katılımcı performansına dayalı önceden tanımlanmış başarı eşiklerine göre içeriğin zorluk düzeyini dinamik olarak uyarlayan kural tabanlı bir algoritma ile çalışmıştır. Kontrol grubuna aynı içerik, geleneksel sınıf tabanlı yöntemle sunulmuştur. Veri toplama araçları olarak Radyasyon Güvenliği Bilgi Testi, Tutum Ölçeği ve Koruyucu Davranış Değerlendirme Formu kullanılmıştır. Veriler bağımlı ve bağımsız örneklem t-testi, ANCOVA ve çoklu doğrusal regresyon analizleri ile değerlendirilmiştir. Uyarlanabilir sistem, önceden tanımlanmış kural tabanlı performans eşiklerine göre çalışmış olup, makine öğrenmesine dayalı özerk model eğitiminden ziyade kural tabanlı uyarlanabilir karar mantığı ile işletilmiştir.

Bulgular: Müdahale sonrası deney grubunun bilgi, tutum ve koruyucu davranış puanlarında kontrol grubuna göre istatistiksel olarak anlamlı bir artış gözlenmiştir ($p<0{,}001$). ANCOVA sonuçları, eğitim türünün final test yeterlilik puanları üzerinde anlamlı ve büyük etkisi olduğunu göstermiştir. Özellikle koruyucu davranış boyutunda önemli bir gelişme gözlenmiştir.

Sonuç: YZ destekli uyarlanabilir eğitim modeli, sağlık profesyonellerinin radyasyon güvenliği yeterliliklerini artırmak için umut verici bir yaklaşım sunmaktadır. Bulgular, performansa dayalı uyarlanabilir içerik ve bireyselleştirilmiş geri bildirimin radyasyon güvenliği eğitiminde sürdürülebilir davranış değişikliğini destekleyebileceğini önermektedir. Modelin genellenebilirliğinin daha büyük örneklem büyüklükleri ve uzun süreli çalışmalarla değerlendirilmesi önerilmektedir.

Anahtar Sözcükler: Radyasyon güvenliği, yapay zeka, sağlık profesyonelleri, uyarlanabilir eğitim, yarı deneysel çalışma.

Introduction

Ionizing radiation has become an indispensable component of diagnostic and treatment processes in modern healthcare. The increasing use of technology, particularly in interventional radiology, cardiology, nuclear medicine, and radiotherapy applications, brings with it the risk of radiation exposure for healthcare workers. The International Atomic Energy Agency (IAEA, 2018) emphasizes that occupational exposure of healthcare personnel can be controlled through appropriate training and safety culture[1]. Similarly, the European Commission (2014) defines continuous and systematic radiation safety training as a mandatory requirement[2].

Radiation safety is not a concept limited solely to technical knowledge. It encompasses a multi-dimensional competency area including knowledge, attitudes, behaviors, and safety culture[3]. Literature reports that the radiation safety knowledge levels of healthcare workers are often insufficient, and in-service training is limited in creating behavioral change[4]. Traditional educational models are generally based on standardized presentations, failing to consider individual learning differences and offering adaptive content based on participant performance[5].

In recent years, artificial intelligence (AI)-based learning systems have offered innovative approaches in health education. Thanks to adaptive learning algorithms, machine learning-based performance analysis, and simulation-supported applications, the learning process can be personalized and real-time feedback can be provided[6]. AI-assisted models in health education are reported to have positive effects on knowledge acquisition and clinical decision-making skills[7,8]. However, experimental studies examining the effectiveness of AI-based educational models specifically in radiation safety are limited. Literature from 2023 onwards reports that AI-assisted adaptive learning systems in health education make significant contributions to clinical decision-making, procedural skill acquisition, and behavioral adjustment[9]. In particular, performance-based feedback and personalized content delivery are highlighted as increasing learning retention.

Measuring the effectiveness of educational interventions in the field of radiation safety should be evaluated not only by the increase in knowledge level but also by attitude change and the development of protective behaviors[10,11]. In this context, it is thought that AI-assisted educational models can support behavioral transformation through adaptive content presentation and simulation-based applications. However, research that demonstrates this assumption with experimental data is needed.

In this study, the effect of an AI-assisted adaptive training model on the radiation safety competency of healthcare professionals was investigated using a quasi-experimental design. The research is expected to contribute to the literature by providing scientific evidence for digital transformation approaches in radiation safety training and developing a model proposal for implementation.

Material and Methods

Research Design

This study is a quasi-experimental pre-test–post-test control group study conducted to evaluate the effect of an artificial intelligence (AI)-based adaptive training model on the radiation safety competence of healthcare professionals. Two parallel groups were included in the study: the experimental group (AI-assisted training) and the control group (traditional training). The intervention period was planned as four weeks.

In the study design, within-group and between-group changes were compared by performing pre- and post-intervention measurements.

Research Population and Sample

The research population consisted of healthcare personnel working with radiation in a university hospital (radiology technicians, interventional radiology nurses, operating room personnel, and nuclear medicine personnel).

The sample size was calculated using the G*Power 3.1 program. With a moderate effect size (d=0.50), a 95% confidence level, and an 80% power assumption, a minimum of 102 participants was required. Considering potential data loss, a total of 120 healthcare professionals were included in the study. Participants were divided into two groups using a simple random assignment method:

• Experimental group (n=60)

• Control group (n=60)

Inclusion criteria:

• Being a healthcare professional working with radiation for at least 1 year

• Voluntary participation

• Not having received radiation safety training in the last 6 months

Exclusion criteria:

• Not regularly attending the training process

• Missing pre-test or post-test data

Group assignment was performed by an independent researcher using a computer-assisted random number generation method. Assignment confidentiality was ensured using the sealed envelope method.

Although participants were randomly assigned to the experimental and control groups, this study was classified as quasi-experimental because participants were recruited using convenience sampling from a single university hospital rather than being randomly selected from the target population. Therefore, randomization was applied only at the group allocation stage.

Data Collection Tools

Radiation Safety Knowledge Test: This is a 25-item multiple-choice test developed by the researchers based on the literature on radiation safety and occupational safety training (IAEA, 2018; European Commission, 2014). Each correct answer is worth 1 point, and an incorrect answer is worth 0 points. The total score ranges from 0 to 25. In the pilot application, the KR-20 reliability coefficient was calculated as 0.84. The content validity of the knowledge test was evaluated by taking into account the opinions of five academics specializing in radiation safety and medical education. The Content Validity Index (CVI) was calculated as 0.89. The test was piloted on 20 healthcare professionals, and its final version was determined by examining the item discrimination indices.

Radiation Safety Attitude Scale: The Radiation Safety Attitude Scale used in this study was developed by the researchers based on the literature on radiation safety and occupational safety training (IAEA, 2018; European Commission, 2014). It consists of 20 items on a 5-point Likert scale (1 = strongly disagree – 5 = strongly agree). The total scale score ranges from 20 to 100. In this study, the Cronbach alpha coefficient was found to be 0.88. As a result of the exploratory factor analysis, it was determined that the scale has a unidimensional structure and explains 61% of the total variance. The Kaiser-Meyer-Olkin (KMO) value was found to be 0.86, and the Bartlett sphericity test was significant ($p<0.001$).

Protective Behavior Assessment Form. The Protective Behavior Assessment Form was developed by the researchers based on international radiation safety guidelines and the ALARA principle (IAEA, 2018). This is a 12-item observational assessment form including personal dosimeter use, lead apron preferences, distance optimization, and compliance with radiation protection practices. Behaviors are coded as 0 (inappropriate) and 1 (appropriate).

Intervention Process

Experimental Group: AI- Assisted Training Model

The experimental group was given an AI-assisted adaptive training program consisting of a total of 8 modules, with two sessions per week for four weeks. The training model consists of the following components:

- Adaptive content guidance based on prior knowledge
- Scenario-based radiation exposure simulations
- Real-time performance analysis
- Personalized feedback reports
- Gamified dose management module

The system automatically adjusted the content difficulty level according to participant performance.

Technical Structure of the Artificial Intelligence System

The system used in this study does not include a machine learning-based predictive model. Instead, a rule-based adaptive decision algorithm is used that dynamically

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

updates the content difficulty level based on participant performance thresholds. In the literature, such systems are defined as rule-based adaptive learning systems evaluated within the scope of symbolic artificial intelligence. Although the system does not perform autonomous data training, it provides performance-based real-time content adaptation and personalized feedback generation.

Control Group: Traditional Training

The control group received standard radiation safety training in a classroom setting with presentations and visual materials over the same period. The training content covered legislation, dose limits, and the use of protective equipment.

Data Collection

The pre-test was administered before the intervention, and the post-test was administered one week after the intervention was completed. The data collection process lasted approximately six weeks. Evaluators who conducted behavioral observations were not blind to group assignment. This may create a risk of observer bias. However, observations were performed using standardized checklists, and evaluation criteria were predefined. The evaluators conducting the behavioral observations received training on standardized evaluation criteria prior to the observations. The observation form items were structured with detailed operational descriptions. Inter-rater agreement was analyzed using Cohen's Kappa coefficient and found to be 0.82. This value indicates a high level of observer agreement.

Ethical Statement

The ethics committee approval was obtained from the Clinical Research Ethics Committee of Istanbul Okan University (Date: 05/02/2025, No: 2025-185). Administrative permission to conduct the study was also obtained from the management of the university hospital where the data were collected prior to the commencement of the study. The purpose of the study was explained to the participants, and their written consent was obtained. Data were collected anonymously and used only for scientific purposes.

Data Analysis

Data were analyzed using SPSS 26.0 software.

- Normality distribution was assessed using the Shapiro–Wilk test.

- Dependent samples t-test was used for within-group comparisons.

- Independent samples t-test was applied for between-group comparisons.

- ANCOVA analysis was performed for post-test results, controlling for pre-test scores as a covariance variable.

- Effect size was calculated using Cohen's d coefficient.
- The significance level was accepted as $p<0.05$.

The percentage improvement for each outcome variable was calculated using the formula: ((post-test score – pre-test score) / pre-test score) × 100.

Results

Demographic Characteristics of Participants

A total of 120 healthcare professionals participated in the study. The demographic characteristics of the experimental and control groups are presented in Table 1.

Table 1. Distribution of participants' demographic characteristics by groups

Variable	Experiment (n=60)	Control (n=60)	Total (n=120)	p
Age (Mean±SD)	32,8±5,6	33,4±6,1	33,1±5,8	0,58
Gender				
Female	36 (%60)	34 (%56,7)	70 (%58,3)	0,64
Male	24 (%40)	26 (%43,3)	50 (%41,7)	
Occupation				
Radiology Technician	21 (%35)	21 (%35)	42 (%35)	0,71
Nurse	17 (%28,3)	17 (%28,3)	34 (%28,3)	
Operating Room Staff	13 (%21,7)	14 (%23,3)	27 (%22,5)	
Nuclear Medicine	9 (%15)	8 (%13,4)	17 (%14,2)	

No statistically significant differences were found between the groups in terms of age, gender, and occupational distribution ($p>0.05$). This indicates that the groups were initially homogeneous.

Pre-Test Comparisons

No significant differences were found between the knowledge, attitude, and behavior scores of the groups before the intervention ($p>0.05$). This finding indicates that the initial levels were similar and that the effect of the intervention can be evaluated more effectively.

Effect on Knowledge Level

The pre-test–post-test knowledge scores of the experimental and control groups are given in Table 2.

Table 2. Pre-test–post-test comparison of knowledge scores

Group	Pretest (Mean±SD)	Posttest (Mean±SD)	t	p	Cohen's d
Experiment	63,1±8,2	88,7±5,9	-20,14	<0,001	1,92
Control	63,8±7,7	77,9±7,4	-9,62	<0,001	1,05

An increase of 40.6% was observed in the knowledge score in the experimental group. The increase in the control group remained at 22.1%. In the post-test comparison between the groups, the knowledge score of the experimental group was significantly higher ($t=8.34$; $p<0.001$). The effect size ($d=1.38$) is very high.

When the pre-test scores were controlled in the ANCOVA analysis, the effect of the training model on the knowledge score was found to be significant ($F=72.51$; $p<0.001$; $\eta^2=0.38$). This value indicates a large effect size.

Effect on Attitude

Table 3. Pre-test–post-test comparison of attitude scores

Group	Pretest (Mean±SD)	Posttest (Mean±SD)	t	p	Cohen's d
Experiment	72,1±6,4	91,2±4,9	-17,83	<0,001	1,74
Control	71,4±6,9	80,8±6,7	-8,41	<0,001	0,93

The increase in attitude scores in the experimental group was 26.5%. ANCOVA results showed a significant difference between groups ($F=44.36$; $p<0.001$; $\eta^2=0.27$).

Effect on Protective Behavior

Table 4. Comparison of protective behavior scores

Group	**Pretest (Mean±SD)**	**Posttest (Mean±SD)**	**Increase (%)**	**p**	**Cohen's d**
Experiment	6,9±1,8	11,2±0,9	62,30%	<0,001	2,01
Control	6,7±1,9	8,5±1,6	26,80%	<0,001	0,89

The effect size in the behavioral dimension (d=2.01) is very high. This result shows that AI-assisted simulation-based training has a strong effect on behavioral change. No clustering at maximum value was observed in the final test score distribution. This indicates that the high effect size obtained is not due to a ceiling effect.

Multiple Regression Analysis Results

A multiple linear regression analysis was applied to determine the variables predicting the total competency score in the radiation safety final test. In the model, the independent variables included training type (0=Traditional, 1=AI-assisted), age, gender (0=Male, 1=Female), and professional experience duration (years). Assumptions were tested before the regression analysis.

• No multicollinearity problem was found (VIF values ranged from 1.08–1.34).

• The Durbin–Watson coefficient was calculated as 1.91, and no autocorrelation was determined.

• It was determined that the residuals showed a normal distribution.

The model was found to be statistically significant ($F(4,115)=21.04$; $p<0.001$) and explains 42% of the variance in the dependent variable ($R^2=0.42$; Adjusted $R^2=0.40$).

Table 5. Variables predicting radiation safety post-test competency score

Variable	B	SH	β	t	p
Fixed	54,27	3,84	—	14,13	<0,001
Education Type	8,96	1,12	0,61	7,98	<0,001
Professional Experience	0,18	0,09	0,14	1,99	0,049
Age	0,07	0,11	0,05	0,64	0,52
Gender	0,84	0,96	0,06	0,87	0,38

According to the analysis results, the type of training is the strongest and most significant predictor of the total competency score ($\beta=0.61$; $p<0.001$). Participants who received AI-assisted training had an average of 8.96 points higher final test competency scores compared to those who received traditional training.

The duration of professional experience is a weak but significant predictor ($\beta=0.14$; $p=0.049$). It was determined that age and gender variables did not significantly contribute to the model ($p>0.05$).

When the standardized coefficients are examined, it is seen that the effect of the type of training is significantly higher than all other variables. No clustering was observed at the maximum values in the final test scores. This indicates that the high effect size obtained is not due to the ceiling effect.

Discussion

The findings showed that the rule-based AI-assisted adaptive training model had a significant positive impact on the radiation safety knowledge and application skills of healthcare professionals. This result is consistent with previous literature highlighting the supportive role of educational technologies in clinical training processes. In particular, adaptive learning approaches have been reported in previous systematic reviews to increase learning outcomes by providing personalized content based on learners' performance[8]. These reviews have shown that adaptive digital learning models have a moderate to high impact on cognitive gain and learning retention compared to traditional in-service training programs. Current meta-analytic findings show that adaptive digital learning models for healthcare professionals have a higher behavioral transfer effect compared to traditional training[9].

In recent years, research examining the contribution of AI-assisted training approaches to the skill development of healthcare professionals has increased. For example, Woolf (2021) stated that systems providing performance-based feedback in health education increased learning motivation and application competence[6]. Similarly, findings indicating that simulation-based training has positive effects on the clinical practice performance of healthcare professionals are widely reported in the literature[11]. In this context, the high effect sizes observed in our current study suggest that the combined use of rule-based adaptive artificial intelligence and simulation-based learning may have strengthened behavioral outcomes.

International studies on radiation safety have reported that while improvements in healthcare professionals' knowledge levels based on traditional training have been observed, the sustainability of behavioral change remains limited[3,10]. These studies show that training interventions aimed solely at increasing knowledge in the field of radiation safety struggle to change behavioral practices. In our study, however, the significant increases observed in behavioral competencies in addition to knowledge acquisition point to the potential of adaptive learning systems to support both cognitive and behavioral outcomes simultaneously. The high effect size obtained in the behavioral dimension ($d>2.00$) may be related to the intensive and short-term application of the intervention. Furthermore, because only short-term measurements were taken, long-term behavioral sustainability was not evaluated. Therefore, the persistence of the effect should be tested with future long-term follow-up studies. However, the study has some limitations, such as its design, population structure, and short-term measurements. The fact that the research was conducted in a single center and that observers were not blinded to group assignment should be considered in terms of the generalizability of the findings and the objectivity of the evaluation.

In addition, the digital literacy levels of the participants were not controlled for, and this should be examined as a variable affecting the benefit obtained from digital education systems. The high effect sizes, especially in behavioral change, may be due to the intensive nature of the intervention and the novelty effect (Hawthorne effect); therefore, it is important to evaluate the findings with long-term follow-up.

Overall, this study provides significant evidence supporting the feasibility and effectiveness of using rule-based adaptive artificial intelligence applications in the development of radiation safety education programs. Supporting future studies with multi-center and long-term follow-up designs will more clearly reveal the sustainable effects of these training models.

Conclusion And Recommendations

This study demonstrated that a rule-based, AI-assisted adaptive training model significantly increased the radiation safety competence of healthcare professionals in terms of knowledge, attitudes, and protective behaviors. The strong effect, particularly in the behavioral dimension, suggests that simulation-based and performance-based feedback-incorporated adaptive training approaches can be effective in developing a safety culture.

The findings indicate that radiation safety training should not be limited to traditional knowledge transfer-based models, and that integrating individualized digital learning approaches into training processes could be beneficial. However, multi-center and follow-up studies are needed to evaluate the applicability and long-term effects of the model in different institutions.

In conclusion, AI-assisted adaptive training systems offer a viable and promising approach to improving the radiation safety competencies of healthcare professionals.

Competing interests: The authors declare that they have no competing interests.

Funding: No financial or material support of any kind was received for the work described in this article.

REFERENCES

1. International Atomic Energy Agency (IAEA). *Radiation protection and safety of radiation sources: International basic safety standards (General Safety Requirements Part 3)*. Vienna: IAEA; 2018.

2. European Commission. *Radiation Protection No. 175: Guidelines on radiation protection education and training of medical professionals*. Luxembourg: Publications Office of the European Union; 2014.

3. Rehani MM, Ciraj-Bjelac O, Vañó E, et al. Radiological protection in fluoroscopically guided procedures performed outside the imaging department. *Radiation Protection Dosimetry*. 2017;165(1–4):510-514. doi: 10.1093/rpd/ncv034.

4. Paolicchi F, Miniati F, Bastiani L, et al. Assessment of radiation protection awareness and knowledge about radiological examination doses among healthcare workers. *Insights into Imaging*. 2016;7(2):233-242. doi: 10.1007/s13244-015-0445-6.

5. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: A meta-analysis. *JAMA*. 2008;300(10):1181-1196.

6. Woolf BP. *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Burlington, MA: Morgan Kaufmann; 2021.

7. Chen L, Chen P, Lin Z. Artificial intelligence in education: A review. *IEEE Access*. 2020;8:75264-75278. doi: 10.1109/ACCESS.2020.2988510.

8. Martin F, Chen Y, Moore RL, Westine CD. Systematic review of adaptive learning research designs, context, strategies and technologies from 2009 to 2018.

Educational Technology Research and Development. 2020;68(4):1903-1929. doi: 10.1007/s11423-020-09793-2.

9. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Artificial intelligence in higher education: A systematic review of research trends. *International Journal of Educational Technology in Higher Education*. 2019;16:39. doi: 10.1186/s41239-019-0171-0.

10. Vañó E, Miller DL, Martin CJ, et al. ICRP Publication 135: Diagnostic reference levels in medical imaging. *Annals of the ICRP*. 2018;46(1):1-144. doi: 10.1177/0146645317717209.

11. Lateef F. Simulation-based learning: Just like the real thing. *Journal of Emergencies, Trauma and Shock*. 2010;3(4):348-352. doi: 10.4103/0974-2700.70743.

12. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*. 2007;39(2):175-191.

13. IBM Corp. *IBM SPSS Statistics for Windows, Version 26.0*. Armonk, NY: IBM Corp; 2019.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 135-152.

Deep Learning and CAD Systems in Breast Cancer Detection A Systematic Review of Diagnostic Accuracy and Clinical Integration

Şeyma Betül SAMSA*, Hina ZAHOOR**

Abstract

Breast cancer continues to be a significant health challenge for women worldwide, and early detection is essential for improving survival outcomes. Over the years, Computer-Aided Diagnosis (CAD) systems have supported radiologists by highlighting suspicious regions in mammograms, ultrasound images, and magnetic resonance imaging (MRI). However, traditional CAD systems evaluate image features based on predefined rules and are associated with high false-positive rates, limiting their ability to adapt to new or complex cases. This study aimed to examine the role of deep learning (DL) in the development of computer-aided diagnosis (CAD) systems and to evaluate the diagnostic accuracy and clinical integration of AI-based approaches in breast cancer detection. A systematic review was conducted by analyzing 13 studies published between 2000 and 2024. Relevant studies were identified through a comprehensive search of the PubMed, IEEE Xplore, and Scopus databases. The included studies were evaluated according to their methodology, deep learning architectures including Convolutional Neural Networks (CNNs), ResNet, EfficientNet, and Vision Transformers and their reported diagnostic performance in terms of accuracy, sensitivity, specificity, and false-negative reduction. The findings were synthesized narratively to compare the advantages and limitations of AI-based diagnostic systems with traditional CAD systems. The findings demonstrate that DL-based models outperform traditional CAD systems. These models contribute to earlier diagnosis by detecting subtle lesions, microcalcifications, and asymmetrical structures that may be overlooked by the human eye. However, several challenges remain, including the availability of large, high-quality datasets, the interpretability of AI decisions, ethical considerations, and integration into routine clinical practice. Deep learning-enhanced CAD systems represent a promising advancement in breast cancer diagnosis by improving diagnostic performance and supporting clinical decision-making. However, their successful implementation depends on high-quality data, transparent and interpretable models, and appropriate ethical and regulatory frameworks to ensure that AI serves as a supportive tool that complements, rather than replaces, radiologists.

Keywords: Deep learning, CAD systems, breast cancer, artificial intelligence, clinical application.

Meme Kanseri Tanısında Derin Öğrenme ve CAD Sistemleri: Tanısal Doğruluk ve Klinik Entegrasyonun Sistematik İncelemesi

Öz

Meme kanseri, dünya çapında kadınlar için önemli bir sağlık sorunu olmaya devam etmekte olup, erken teşhis sağkalım sonuçlarının iyileştirilmesi açısından kritik öneme sahiptir. Yıllar boyunca Bilgisayar Destekli Tanı (CAD) sistemleri, mamografi, ultrason ve manyetik rezonans görüntüleme (MRG) incelemelerinde şüpheli bölgeleri belirleyerek radyologlara destek olmuştur. Bununla birlikte, geleneksel

Derleme Makale (Review Article)
Geliş / Received: 31.12.2025 & **Kabul / Accepted:** 20.07.2026
DOI: https://doi.org/10.38079/igusabder.1852670
* Student, Istanbul Gelisim University, Faculty of Health Sciences, Department of Nursing, Istanbul, Türkiye.
E-mail: seyma.seymasamsa@gmail.com ORCID https://orcid.org/0009-0005-0793-8266
** Asst. Prof., Istanbul Gelisim University, Istanbul, Türkiye. E-mail: hzahoor@gelisim.edu.tr
ORCID https://orcid.org/0000-0003-2322-5678

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

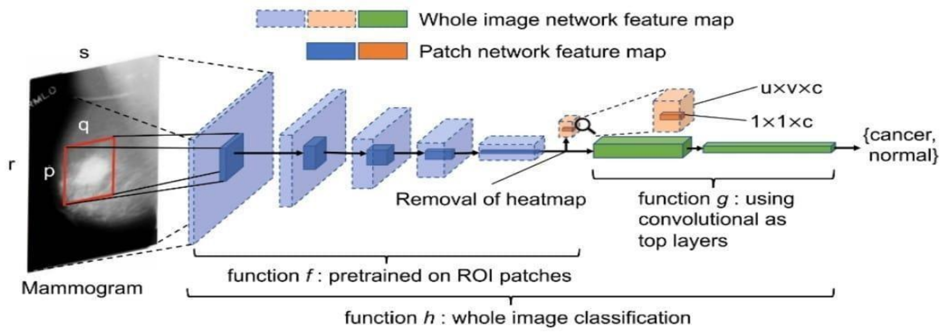
CAD sistemleri görüntü özelliklerini önceden tanımlanmış kurallara göre değerlendirmekte ve yüksek yanlış pozitif oranları nedeniyle yeni veya karmaşık olgulara uyum sağlamada sınırlılıklar göstermektedir. Bu çalışmanın amacı, derin öğrenmenin (DL) bilgisayar destekli tanı (CAD) sistemlerinin geliştirilmesindeki rolünü incelemek ve yapay zekâ tabanlı yaklaşımların meme kanseri tanısındaki tanısal doğruluğunu ve klinik entegrasyonunu değerlendirmektir. PubMed, IEEE Xplore ve Scopus veri tabanlarında 2000–2024 yılları arasında yayımlanan çalışmalar sistematik olarak tarandı. Dahil edilme kriterlerini karşılayan 13 çalışma incelemeye alındı. Çalışmalar, metodolojik özellikleri, kullanılan derin öğrenme mimarileri (Evrişimsel Sinir Ağları, ResNet, EfficientNet ve Vision Transformers) ile doğruluk, duyarlılık, özgüllük ve yanlış negatif oranları açısından değerlendirildi. Bulgular, yapay zekâ tabanlı tanı sistemlerinin geleneksel CAD sistemlerine göre avantajlarını ve sınırlılıklarını karşılaştırmak amacıyla anlatımsal olarak sentezlendi. Bulgular, DL tabanlı modellerin geleneksel CAD sistemlerine kıyasla üstün performans sergilediğini göstermektedir. Bu modeller, radyologlar tarafından gözden kaçabilecek ince lezyonları, mikrokalsifikasyonları ve asimetrik yapıları tespit ederek erken tanıya katkı sağlamaktadır. Bununla birlikte, büyük ve yüksek kaliteli veri setlerine erişim, yapay zekâ kararlarının yorumlanabilirliği, etik konular ve rutin klinik uygulamaya entegrasyon gibi önemli zorluklar devam etmektedir. Derin öğrenme ile güçlendirilmiş bilgisayar destekli tanı (CAD) sistemleri, tanısal performansı artırarak ve klinik karar verme süreçlerini destekleyerek meme kanseri tanısında umut verici bir gelişme sunmaktadır. Bununla birlikte, bu sistemlerin başarılı bir şekilde klinik uygulamaya geçirilmesi; yüksek kaliteli veri kullanımına, şeffaf ve yorumlanabilir modellerin geliştirilmesine ve uygun etik ile düzenleyici çerçevelerin oluşturulmasına bağlıdır. Yapay zekâ, radyologların yerini alan değil, klinik karar vermeyi destekleyen tamamlayıcı bir araç olarak kullanılmalıdır.

Anahtar Sözcükler: Derin öğrenme, CAD sistemleri, meme kanseri, yapay zekâ, klinik uygulama.

Introduction

Breast cancer is one of the most significant health threats to women worldwide, and early and accurate diagnosis is crucial for improving treatment outcomes and increasing survival rates[1,2]. Traditional Computer- Diagnosis (CAD) systems have long assisted radiologists by marking suspicious areas in mammograms, ultrasound, and MRI images, but their dependence on handcrafted features and limited adaptability frequently result in false positives[3,4]. In contrast, modern artificial intelligence approaches, especially deep learning models like Convolutional Neural Networks, ResNet-50, EfficientNet, and Vision Transformers, have enabled this leap by learning complex hierarchical patterns directly from raw medical images[6-8]. They also demonstrated much higher sensitivity and diagnostic accuracy related to the detection of malignant lesions, including microcalcifications and subtle asymmetries that may be missed by conventional methods[9-11]. Yet, despite their promise, several practical challenges exist for translating DL based systems into routine clinical practice: they require large, well-annotated datasets; their decision processes are often not transparent; and integration into current clinical workflow processes remains an open issue[12,13]. This review systematically examines and compares the diagnostic performance of deep learning (DL) and computer-aided diagnosis (CAD) systems in breast cancer detection, focusing on their accuracy, feasibility, and potential for clinical translation. As illustrated in Figure 1, the DL-based architecture streamlines complex mammography data by extracting key features through a multi-stage process, ultimately providing a definitive classification as either 'cancer' or 'normal'[9].

Figure 1. Performance comparison of DL-based mammography models.



There is a need for more sophisticated diagnostic technologies to increase survival rates. Although mammographic imaging, ultrasonography, and magnetic resonance imaging are established methods of diagnosing cancer in the breasts, these techniques have certain drawbacks, such as diminished sensitivity in dense tissues and high costs[14,15].

Conventional CAD tools help radiologists spot unusual shapes and textural patterns according to known rules. However, such a need for manual feature representation makes them prone to high false positives[3,4]. AI and ML technologies, including CNN, have brought about a paradigm shift in detecting complex features from raw images. CNN architectures, involving convolution and pooling, have shown high accuracy in distinguishing between malignant and benign lesions[16-24].

Many studies have demonstrated the effectiveness of CNNs in breast cancer diagnosis. For example, Liu et al. reported high diagnostic performance in mammogram classification, while Wang et al. demonstrated the effectiveness of CNN-based models for breast lesion classification[16,17]. Present-day applications include AI-assisted mammography systems that have demonstrated promising performance in clinical breast cancer screening[9,20].

Nevertheless, "black box" interpretability and data quality are still considered challenges in the integration of these solutions in the clinical environment[21,25]. There is evidence that although conventional CAD systems are particularly dependent on radiologist assessment regarding diagnosis, DL solutions are more transferable[3,18,21]. Current evidence suggesting improvements made possible by the integration of DL's learning capabilities with CAD's systematic analysis is the reduction of false positives[20].

Material and Methods

This study was conducted as a systematic literature review to evaluate the diagnostic performance, clinical effectiveness, and implementation of deep learning (DL)-based artificial intelligence (AI) systems for breast cancer detection in mammography screening. The review was designed and reported in accordance with the Preferred

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Reporting Items for Systematic Reviews and Meta-Analyses 2020 (PRISMA 2020) Statement.

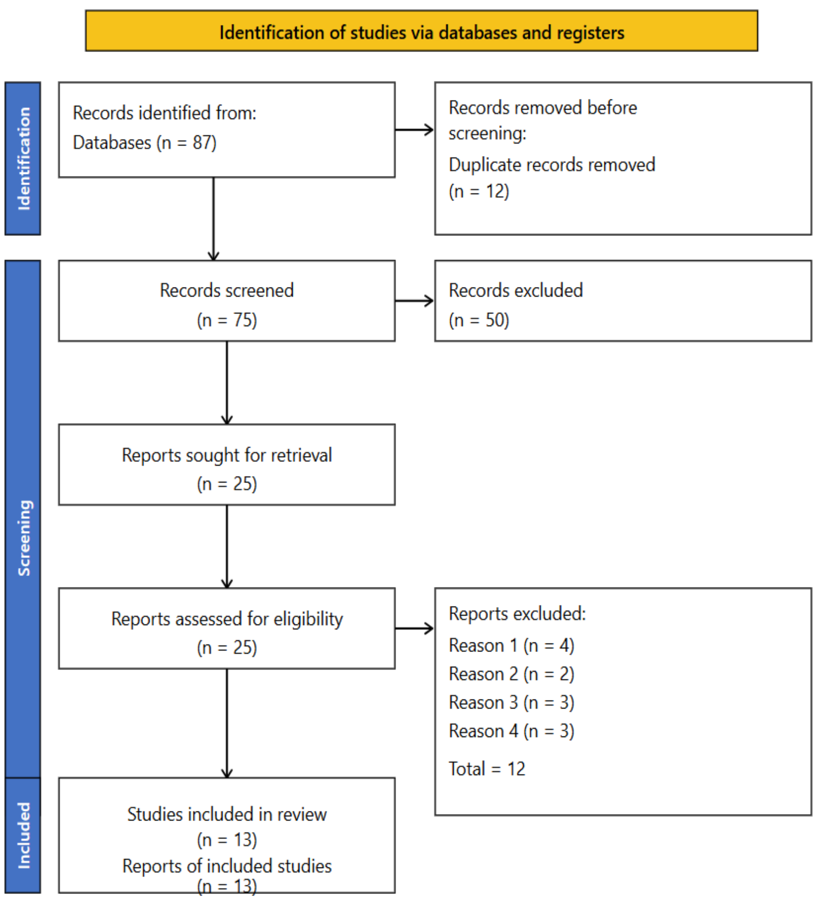
A comprehensive literature search was performed using the PubMed, IEEE Xplore, and Scopus databases to identify relevant studies published between January 2000 and December 2024. The search strategy combined database-specific keywords and Boolean operators (AND/OR) related to breast cancer, mammography, deep learning, artificial intelligence, convolutional neural networks (CNNs), and computer-aided diagnosis (CAD). The complete search strategy, together with the predefined inclusion and exclusion criteria, is presented in Table 1.

Table 1. Literature search strategy and eligibility criteria

Database	Search Strategy	Inclusion Criteria	Exclusion Criteria
PubMed	("breast cancer" OR "breast neoplasm") AND ("deep learning" OR CNN OR "artificial intelligence") AND ("computer-aided diagnosis" OR CAD)	Original peer-reviewed studies; English; 2000–2024; DL-based CAD for breast cancer; reports accuracy, sensitivity, specificity, or AUC	Reviews, editorials, conference abstracts, case reports, non-English, animal studies, no empirical data
IEEE Xplore	("deep learning" AND "breast cancer") AND ("computer-aided diagnosis" OR mammography)	Same criteria as PubMed	Same criteria as PubMed
Scopus	("breast cancer") AND ("deep learning" OR CNN OR ResNet OR EfficientNet OR Vision Transformer) AND ("diagnostic imaging")	Same criteria as PubMed	Same criteria as PubMed

The initial search identified 87 records. After duplicate removal, titles and abstracts were independently screened according to the predefined eligibility criteria. Full-text articles were subsequently assessed for eligibility, and thirteen studies met the inclusion criteria and were included in the final review. The study selection process is illustrated in the PRISMA flow diagram (Figure 2).

Figure 2. PRISMA 2020 flow diagram illustrating the study selection process.



Data extraction was performed using a standardized extraction form to ensure consistency across studies. The following information was collected from each eligible article: first author, publication year, study design, study population, imaging modality, AI model or deep learning architecture, validation strategy, diagnostic performance outcomes (including sensitivity, specificity, accuracy, area under the receiver operating characteristic curve [AUC], cancer detection rate [CDR], recall rate [RR], positive predictive value [PPV], and workload reduction where available), and the principal findings. Due to substantial heterogeneity in study populations, AI models, imaging modalities, validation strategies, and reported outcome measures, a quantitative meta-analysis was not considered appropriate. Therefore, the findings were synthesized using a qualitative narrative approach.

The methodological quality and risk of bias of the included diagnostic accuracy studies were evaluated using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. Four methodological domains were assessed: Patient Selection, Index Test, Reference Standard, and Flow and Timing. Each domain was classified as having a low, high, or unclear risk of bias according to the QUADAS-2 guidance. The assessment focused on methodological quality and reporting transparency rather than the reported diagnostic performance of the AI systems. The results of the quality assessment are presented in Table 2.

Table 2. QUADAS-2 Risk of Bias Assessment of the Included Diagnostic Accuracy Studies

Study	Patient Selection	Index Test	Reference Standard	Flow & Timing	Overall Risk
Sung et al. (2021)[1]	Low Risk	Low Risk	Low Risk	Unclear Risk	Low Risk
Rodríguez-Ruiz et al. (2019)[3]	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Dang et al. (2022)[4]	Moderate Risk	Low Risk	Moderate Risk	Low Risk	Moderate Risk
McKinney et al. (2020)[9]	Low Risk	Unclear Risk	Low Risk	Low Risk	Low Risk
Ribli et al. (2018)[10]	High Risk	Unclear Risk	Low Risk	Unclear Risk	High Risk
Schaffter et al. (2020)[12]	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Yala et al. (2019)[13]	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Lauritzen et al. (2022)[19]	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Lotter et al. (2021)[23]	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Lauritzen et al. (2024)[15]	Low Risk	Low Risk	Low Risk	Moderate Risk	Low Risk
Lång et al. (2023)[5]	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Elhakim et al. (2024)[26]	Low Risk	Low Risk	Low Risk	Moderate Risk	Low Risk
Kim et al. (2020)[27]	High Risk	Low Risk	Low Risk	Low Risk	Moderate Risk

For the Patient Selection domain, studies were considered to have a low risk of bias when they included representative screening populations, clearly defined eligibility criteria, and avoided inappropriate exclusions. Studies using cancer-enriched datasets, retrospective convenience sampling, or non-consecutive participant selection were considered to have an increased risk of selection bias, particularly when these characteristics were insufficiently justified or likely to reduce the representativeness of the screening population.

The Index Test domain evaluated whether AI predictions were interpreted independently of the reference standard and whether diagnostic thresholds were predefined before model evaluation. Studies that used prespecified AI decision thresholds and interpreted the index test without knowledge of the reference standard were considered to have a low risk of bias. When blinding procedures or threshold prespecification were not clearly reported, the domain was classified as having an unclear risk of bias.

The Reference Standard domain assessed whether breast cancer status was established using an appropriate reference standard, including histopathological confirmation for positive cases and adequate imaging or clinical follow-up for negative examinations. Studies fulfilling these criteria were considered to have a low risk of bias in this domain.

The Flow and Timing domain examined whether all participants received an appropriate reference standard, whether the interval between the index test and disease verification was acceptable, and whether all eligible participants were included in the final analysis. Prospective studies with complete follow-up and consistent application of the reference standard were considered to have a low risk of bias. Studies with incomplete reporting of participant flow, verification procedures, or follow-up of negative cases were classified as having an unclear risk of bias.

Overall, the QUADAS-2 assessment was undertaken to facilitate interpretation of the reported diagnostic performance, identify potential sources of methodological bias, and strengthen the transparency and reliability of the systematic review.

As this review was based exclusively on previously published studies and did not involve human participants or identifiable patient data, ethical approval and informed consent were not required.

Table 3. Literature Review Synthesis Matrix Entry

Author/ Year	**Purpose**	**Methods**	**Results**	**Key ideas**	**Strengths**	**Notes/my conclusions**
Sung et al. (2021)[1]	To provide global cancer statistics, including breast cancer incidence, mortality, and trends.	Epidemiological study using data from GLOBOCAN 2020.	Breast cancer became the most commonly diagnosed cancer worldwide, with regional variation in incidence and mortality.	Global cancer trends; Breast cancer epidemiology; Public health implications	Strength: Large-scale, up-to-date data source. Limitation: May underreport in countries with poor data infrastructure. Further research: Need for real-time cancer registries.	Highlights the urgency of effective diagnostic tools like AI to reduce global burden.
Rodríguez-Ruiz et al. (2019)[3]	To compare the diagnostic performance of radiologists interpreting screening mammograms with and without artificial intelligence (AI) assistance and to determine whether AI improves breast cancer detection.	Retrospective, fully crossed, multi-reader multi-case diagnostic accuracy study including 240 digital screening mammograms interpreted by 14 MQSA-qualified radiologists. Each examination was read twice (unaided and AI-assisted). Diagnostic performance was assessed using AUC, sensitivity, specificity, and reading time.	AI assistance significantly improved diagnostic performance by increasing AUC (0.89 vs. 0.87) and sensitivity (86% vs. 83%), while specificity showed a slight improvement and reading time remained unchanged. AI alone achieved diagnostic performance comparable to the average radiologist.	AI can function as an effective decision-support tool that enhances radiologists' cancer detection without increasing interpretation time. The study supports integrating AI into routine mammography screening as a second-reader system.	Multi-reader multi-case design; direct comparison of AI-assisted and unaided interpretation; clinically relevant performance metrics (AUC, sensitivity, specificity); evaluation of workflow efficiency through reading time analysis.	This study provides strong evidence that AI support improves radiologists' diagnostic accuracy while maintaining workflow efficiency. Its robust diagnostic accuracy design makes it highly suitable for QUADAS-2 quality assessment and supports the clinical value of AI-assisted mammography screening.
Dang et al. (2022)[4]	To evaluate whether radiologists assisted by an artificial intelligence (AI) system can improve BI-RADS classification accuracy, increase breast cancer detection performance, and maintain efficient mammogram interpretation time during screening mammography.	A retrospective multi-reader, multi-case crossover study was conducted using 314 screening mammograms collected between 2012 and 2020 from a French screening program. Twelve radiologists (eight senior and four junior) independently interpreted mammograms in two reading sessions separated by a 4-week washout period, with and without AI assistance (Mammoscreen™ v1.2). Reading order was randomized, and each radiologist assigned BI-RADS categories, lesion locations, and continuous suspicion scores. Primary outcomes included inter-reader agreement (quadratic Cohen's kappa), diagnostic accuracy (AUC), and interpretation time.	AI assistance significantly improved radiologists' agreement in BI-RADS classification (κ increased from 0.549 to 0.626) and increased diagnostic accuracy (AUC improved from 0.74 to 0.77, *p*=0.004). Reading time was not significantly affected (106 s without AI vs. 102 s with AI, *p*=0.754), indicating that AI enhanced performance without reducing efficiency.	This study demonstrates that AI can function as an effective decision-support tool for radiologists by improving the consistency and accuracy of BI-RADS categorization without increasing interpretation time. Rather than replacing radiologists, the AI system enhanced clinical decision-making by providing lesion localization and malignancy scores, leading to more reliable screening assessments. The findings support the integration of AI into routine mammography workflows to improve diagnostic quality while preserving clinical efficiency.	A major strength of the study is its multi-reader, multi-case crossover design, which minimizes inter-reader variability and allows each radiologist to act as their own control. The inclusion of both experienced and junior radiologists improves the generalizability of the findings across different expertise levels. Additionally, the use of an FDA-approved and CE-marked AI system, randomized reading order, washout period, and objective performance metrics (kappa coefficient, AUC, and reading time) provides strong methodological rigor.	This study provides strong evidence that AI is most valuable as a clinical decision-support tool rather than as a replacement for radiologists. By improving BI-RADS classification consistency and diagnostic performance without increasing reading time, AI has the potential to reduce inter-observer variability and standardize breast cancer screening. However, because the study was conducted in a controlled retrospective reader-study setting with an enriched cancer dataset, prospective real-world validation is still needed before widespread

						clinical implementation.
McKinney et al. (2020)[9]	To assess whether AI using deep learning can outperform traditional CAD methods and radiologists in breast cancer diagnosis.	CNN-based model by Google Health tested on a large mammography dataset; compared with multiple radiologists.	AI significantly reduced false positives and false negatives. Best outcomes achieved when AI worked with radiologists.	AI integration in clinical practice; deep learning in diagnostics; collaborative diagnosis.	Strength: High accuracy and potential to reduce unnecessary biopsies. Limitation: Needs further validation on larger cohorts and regulatory approval. Future research: Broader clinical validation.	Highlights implementation potential and raises questions on regulation, clinician training, and system integration in real-world settings.
Ribli et al. (2018)[10]	To develop and evaluate a deep learning-based model for detecting breast cancer in mammography images, comparing it with traditional CAD systems.	Model development and performance evaluation study using mammography datasets; AI-based image analysis.	The model achieved 92% sensitivity and 87% specificity, outperforming traditional CAD systems in accuracy.	AI-based CAD systems; Deep learning performance; Diagnostic accuracy improvement	Strength: High accuracy results with potential clinical impact. Limitation: Limited by dataset and lack of integration testing in clinical workflow. Further research: Validation in real clinical settings and broader datasets.	Promising technical performance, but requires further testing for integration into clinical decision-making processes.
Schaffter et al. (2020)[12]	To evaluate whether deep learning algorithms alone and in combination with radiologists improve the diagnostic accuracy of screening mammography.	Multicenter retrospective diagnostic accuracy study using two independent screening mammography datasets from the United States and Sweden. AI algorithms were developed through the Digital Mammography DREAM Challenge and validated on external datasets. Performance was compared with radiologists using biopsy-confirmed outcomes, ROC analysis, sensitivity, specificity, and AUC.	The best individual AI algorithm achieved high diagnostic performance but did not consistently outperform radiologists. However, combining AI with radiologist assessment significantly improved specificity while maintaining high sensitivity.	AI should function as a clinical decision-support tool rather than replacing radiologists. External validation and integration with clinician interpretation improve screening performance.	Large multicenter datasets, independent external validation, biopsy-confirmed reference standard, standardized diagnostic accuracy evaluation, comparison with radiologists, STARD reporting guideline.	This study demonstrates that combining AI with radiologists provides better screening performance than either alone. Its robust multicenter design and validated diagnostic accuracy methodology make it highly suitable for evidence synthesis and QUADAS-2 quality assessment.
Yala et al. (2019)[13]	To develop and validate a deep learning (DL) mammography-based model for predicting 5-year breast cancer risk and compare its performance with established clinical risk models.	Retrospective diagnostic accuracy study including 88,994 consecutive screening mammograms from 39,558 women. Patients were randomly assigned to training, validation, and independent test sets. Cancer outcomes were confirmed through tumor registries and pathology records. Model performance was evaluated using ROC analysis and AUC.	The hybrid deep learning model achieved the highest predictive performance (AUC 0.70), outperforming both the traditional Tyrer–Cuzick model (AUC 0.62) and the logistic regression model using conventional risk factors (AUC 0.67).	Deep learning can predict future breast cancer risk directly from mammographic images and improves long-term risk stratification compared with conventional clinical risk models.	Large consecutive screening cohort, independent training/validation/test datasets, tumor registry–confirmed outcomes, external comparison with established risk models, full-field mammography analysis.	This study demonstrates that mammography-based deep learning models can substantially improve breast cancer risk prediction beyond traditional risk assessment tools, supporting the integration of AI into personalized screening strategies.
Lauritzen et al. (2024)[15]	To evaluate the accuracy and feasibility of three AI-	A retrospective simulation study was conducted using 249,402 screening	AI replacing the first reader and AI-based triage reduced	This study systematically compared three different AI	The study analyzed one of the largest population-based mammography	This study provides practical evidence that AI can substantially

	integrated mammography screening strategies that replace one or both radiologists in a double-reading screening program while maintaining cancer detection performance and reducing screening workload.	mammograms from 149,495 women in a population-based breast cancer screening program. A commercially available deep learning AI system (Lunit INSIGHT MMG v1.1.7.1) was evaluated in three screening scenarios: AI replacing the first reader, AI replacing the second reader, and AI-based triage. Diagnostic performance was compared with standard double reading using sensitivity, specificity, cancer detection rate (CDR), PPV, NPV, recall rate, arbitration rate, and AUC.	radiologist workload by approximately 50% while maintaining cancer detection accuracy comparable to standard double reading. AI replacing the second reader also reduced workload but resulted in a small yet statistically significant reduction in sensitivity. The AI-based triage strategy demonstrated the most favorable balance between diagnostic performance and workflow efficiency.	deployment strategies within a real-world mammography screening workflow. The findings demonstrated that the clinical impact of AI depends on how it is integrated into the screening process, with AI-assisted triage offering the best compromise between maintaining diagnostic accuracy and reducing radiologist workload.	screening cohorts and evaluated multiple clinically realistic AI implementation scenarios using a commercially approved deep learning system. Comprehensive assessment of cancer detection, workload, arbitration, and diagnostic performance provides strong evidence to support future AI integration into breast cancer screening programs.	improve the efficiency of double-reading mammography screening without compromising diagnostic accuracy when appropriately integrated into the workflow. Among the evaluated strategies, AI-assisted triage appears to be the most clinically feasible approach. However, because the results are based on retrospective simulations, prospective clinical implementation studies remain necessary before widespread adoption.
Lotter et al. (2021)[23]	To develop and validate an annotation-efficient deep learning model for breast cancer detection in digital mammography and digital breast tomosynthesis (DBT), and to compare its diagnostic performance with expert breast radiologists.	A retrospective multicenter diagnostic accuracy study was conducted using five mammography datasets from the United States, the United Kingdom, and China. A progressively trained deep learning model was developed using both strongly and weakly annotated mammographic images. The model was evaluated on independent external test datasets and compared with five experienced breast-imaging radiologists using ROC analysis, sensitivity, specificity, and AUC.	The AI model outperformed all five breast-imaging specialists, increasing sensitivity by approximately 14% while maintaining high specificity. It achieved excellent diagnostic performance across multiple independent datasets (AUC up to 0.97) and demonstrated strong generalizability for both digital mammography and digital breast tomosynthesis.	This study demonstrated that a progressively trained deep learning framework using both strongly and weakly annotated mammography data can achieve high diagnostic accuracy across multiple populations and imaging modalities. The model maintained lesion localization capability while improving cancer detection performance and showed consistent generalizability on independent external datasets, including digital mammography and digital breast tomosynthesis.	The study included a large, multicenter, and multinational dataset with extensive external validation across different countries, imaging systems, and patient populations. Its robust methodological design, comparison with expert breast radiologists, evaluation on independent test sets, and use of clinically verified reference standards substantially enhanced the reliability, reproducibility, and clinical applicability of the findings.	This study provides strong evidence that deep learning can significantly improve breast cancer detection performance while maintaining robust generalizability across diverse clinical settings. The multicenter design and external validation make it one of the strongest diagnostic accuracy studies included in this review.
Lauritzen et al. (2022)[19]	To evaluate whether an artificial intelligence (AI)-based mammography screening protocol could safely reduce radiologists' workload while maintaining diagnostic performance across different breast density categories.	A retrospective simulation study including 114,421 consecutive screening mammograms from the Danish breast cancer screening program. Mammograms were analyzed using the Transpara AI system and compared with the standard double-reading workflow. Diagnostic performance was assessed using sensitivity, specificity, AUC, workload	AI-based screening achieved non-inferior sensitivity (69.7% vs. 70.8%), significantly higher specificity (98.6% vs. 98.1%), reduced radiologist workload by 62.6%, and decreased false-positive recalls by 25.1%, with consistent performance	This study demonstrated that AI can effectively triage mammography examinations by identifying normal, moderate-risk, and suspicious cases before radiologist review. The proposed AI-assisted workflow maintained diagnostic accuracy while substantially reducing radiologists' workload and	The study used a very large consecutive screening cohort with more than 114,000 women, independent development and validation datasets, a commercially validated AI system, standardized reference standards including interval cancers, and comprehensive statistical analyses. These methodological	This study provides strong evidence that AI can be integrated into routine breast cancer screening to optimize workflow efficiency without compromising diagnostic accuracy, making it highly relevant for future AI-assisted screening strategies.

		reduction, and false-positive rates.	across all BI-RADS breast density categories.	improving screening efficiency across different breast density groups, suggesting that AI can safely support population-based breast cancer screening programs.	strengths enhance the reliability, external validity, and clinical applicability of the findings.	
Lång et al. (2023)[5]	To evaluate whether AI-supported mammography screening can safely reduce radiologists' workload while maintaining cancer detection performance in a population-based randomized screening trial.	80,033 women from four Swedish screening centers (2021–2022); randomized controlled trial.	AI-supported screening achieved a cancer detection rate comparable to standard double reading while reducing radiologist workload by 44.3%. Recall and false-positive rates remained similar, and the positive predictive value of recall improved with AI assistance.	This study demonstrated, in a real-world randomized clinical trial, that AI-supported screening maintained screening safety while substantially reducing radiologist workload. AI triaged examinations according to malignancy risk and optimized the reading workflow without compromising cancer detection.	The study is one of the first large-scale randomized controlled trials evaluating AI in routine breast cancer screening. Its multicenter population-based design, large sample size, standardized workflow, and clinically meaningful workload reduction provide strong evidence supporting AI implementation in screening practice.	The first large randomized clinical trial demonstrated that AI can safely support population-based mammography screening by substantially reducing reading workload without compromising cancer detection. Long-term interval cancer follow-up is still required to confirm clinical effectiveness.
Elhakim et al. (2024)[26]	To evaluate the diagnostic accuracy and feasibility of three artificial intelligence (AI)-integrated mammography screening strategies by assessing whether AI can replace one or both radiologists in a double-reading screening program while maintaining cancer detection performance and reducing radiologist workload.	This retrospective multicenter study analyzed 272,008 consecutive screening mammograms collected in the Region of Southern Denmark between 2014 and 2018. After applying exclusion criteria, 249,402 screening examinations were included. Three AI-integrated screening scenarios (AI replacing the first reader, second reader, or both readers through triage) were simulated using the FDA- and CE-approved Lunit INSIGHT MMG v1.1.7.1 system and compared with standard double reading with arbitration. Cancer outcomes were verified through national breast cancer registries and histopathology. Diagnostic performance was evaluated using sensitivity, specificity, cancer detection rate (CDR), positive predictive value (PPV), negative predictive value (NPV), recall rate,	Replacing the first radiologist or using AI as a triage tool reduced screening workload by approximately 49–50% without compromising cancer detection performance. AI triage achieved the highest overall performance, demonstrating improved sensitivity (+1.33%), higher cancer detection rate, and lower arbitration rate compared with conventional double reading. Replacing the second radiologist also reduced workload but resulted in a small yet statistically significant reduction in sensitivity (−1.58%, $P < .001$).	This study investigated different strategies for integrating AI into population-based mammography screening rather than evaluating AI as an independent diagnostic system. By comparing three implementation scenarios, the authors demonstrated that AI can substantially reduce radiologists' workload while maintaining diagnostic accuracy when deployed appropriately. The findings suggest that AI performs best as either the first reader or as a triage tool that automatically classifies low- and high-risk examinations while reserving radiologist assessment for intermediate-risk cases. The study highlights that the clinical impact of AI depends not only on algorithm performance but also on how it is incorporated into	A major strength of this study is its large multicenter population-based cohort, including nearly 250,000 consecutive screening mammograms, which closely reflects routine clinical practice. The authors compared multiple realistic AI deployment strategies instead of evaluating AI alone, making the findings highly relevant for future implementation in organized breast screening programs. The use of validated national cancer registries, histopathological confirmation, standardized performance metrics, and a commercially available FDA- and CE-approved AI system further strengthens the reliability and clinical applicability of the results.	This study provides practical evidence that the effectiveness of AI in breast cancer screening depends on its integration into the clinical workflow rather than simply replacing radiologists. Among the evaluated strategies, AI-assisted triage offered the best balance between maintaining cancer detection accuracy and substantially reducing reading workload. Although the study was retrospective and based on simulated screening scenarios, its large population and clinically realistic design make it highly relevant for future implementation studies and support the role of AI as a workflow optimization tool rather than a complete

		arbitration rate, and AUC.		the screening workflow.		substitute for human readers.
Kim et al. (2024)[27]	To prospectively evaluate whether AI-assisted computer-aided detection (AI-CAD) improves the diagnostic performance of radiologists in population-based mammography screening, and to compare breast specialists, general radiologists, and standalone AI in terms of cancer detection and recall rates.	A prospective, multicenter population-based study was conducted across six academic hospitals in South Korea. Women aged ≥40 years undergoing routine digital mammography screening were enrolled, while those with previous breast cancer or mammoplasty were excluded. Breast-specialized radiologists first interpreted mammograms without AI and then with AI-CAD assistance using the Lunit INSIGHT MMG system. Standalone AI performance and a simulated reading study involving five general radiologists were also evaluated. The primary outcomes were cancer detection rate (CDR) and recall rate (RR), with secondary analyses including positive predictive value (PPV), subgroup analyses by age and breast density, and comparisons between AI-assisted and non-assisted interpretations. Statistical analyses used generalized estimating equations, logistic regression, and McNemar-based sample size estimation.	AI assistance improved the cancer detection rate of breast-specialized radiologists while maintaining an acceptable recall rate. Standalone AI achieved diagnostic performance comparable to experienced breast radiologists, and AI assistance substantially improved the performance of general radiologists, narrowing the gap between specialists and non-specialists. The benefits of AI were observed across different age groups and breast density categories.	The study provides prospective evidence that AI-CAD can enhance real-world breast cancer screening by increasing cancer detection while preserving clinical decision-making by radiologists. AI particularly benefits less-experienced general radiologists and may help address workforce shortages without replacing specialist oversight.	This study is strengthened by its prospective multicenter design, large population-based cohort, real-world clinical implementation, and evaluation within routine national screening practice. It directly compares specialist radiologists, general radiologists, and standalone AI using standardized outcomes, making the findings highly applicable to clinical implementation.	Unlike retrospective simulation studies, this prospective trial demonstrates the practical clinical value of AI-CAD in routine screening. The sequential reading design reflects real-world workflow and shows that AI functions effectively as a decision-support tool rather than an independent replacement for radiologists. The inclusion of both specialist and general radiologists provides valuable evidence for implementing AI in settings with limited breast imaging expertise.

Results

A total of 87 records were identified through the database search. After duplicate removal and eligibility screening, 13 studies met the inclusion criteria and were included in the final systematic review (Figure 2). The characteristics and principal findings of the included studies are summarized in Table 3.

Overall, the included studies consistently demonstrated that deep learning (DL)-based computer-aided diagnosis (CAD) systems achieved superior diagnostic performance compared with conventional CAD approaches. Convolutional neural network (CNN)-based models were the most frequently investigated architectures and generally reported higher diagnostic accuracy, sensitivity, and specificity across different imaging modalities, particularly mammography and breast ultrasound[9,10,17-19,28].

Several studies reported that DL-assisted systems reduced false-positive and false-negative findings while improving lesion detection and classification. Studies by McKinney et al., Ribli et al., Kim et al., and Wu et al. demonstrated that AI-assisted diagnosis performed comparably to, or better than, experienced radiologists, especially when integrated into the clinical workflow[9,10,20,28]. These findings suggest that human–AI collaboration may improve breast cancer screening performance rather than replace radiologists[12,13,21].

The reviewed studies also highlighted several challenges affecting the clinical implementation of AI-based CAD systems. These included limited external validation, variability in imaging datasets, lack of standardized evaluation protocols, regulatory considerations, and ethical concerns regarding transparency, accountability, and patient data privacy[12,13,21,22,29,30]. Furthermore, several authors emphasized the need for multicenter prospective studies to validate AI models before routine clinical implementation[8,13,25].

Overall, the findings indicate that deep learning-based CAD systems have substantial potential to improve the accuracy and efficiency of breast cancer detection. However, further large-scale clinical validation and standardized evaluation methods are required before widespread adoption in routine clinical practice[9,10,13,28].

Table 3. Comparison of traditional CAD vs. Deep Learning Models

Feature	Deep Learning- Based Systems	Traditional CAD Systems
Accuracy	High (92%-95%)	Moderate (~85%)
Sensitivity	High (90%-95%)	Moderate (80%-85%)
Specificity	Moderate (85%-90%)	High (90%)
False Positive Rate	High (7.5%-15%)	Low (5%-10%)
False Negative Rate	Low (2.5%-5%)	Moderate (8%-12%)
AUC (Area Under Curve)	High (0.90-0.95)	Moderate (0.80-0.85)
Clinical Application	Emerging, under development	Widely used in clinical practice
Data Requirement	Requires large datasets	Works with smaller datasets

Diagnostic Performance of Deep Learning Models

It is evident from the studies included in this systematic review that deep learning (DL) techniques, particularly those based on convolutional neural networks (CNNs), demonstrate higher sensitivity and specificity than traditional computer-aided diagnosis (CAD) systems, thereby reducing false-negative findings commonly associated with rule-based approaches[9,10,25]. ResNet and EfficientNet have shown improved performance in detecting malignant lesions at earlier stages than traditional diagnostic systems[8,9].

Comparison with Traditional CAD Systems

Traditional CAD systems are limited by predefined rules and textural patterns, often leading to high false-positive rates and unnecessary patient stress[3,4]. In contrast, AI-based approaches demonstrate a higher capacity for feature extraction directly from raw medical images, which enhances diagnostic accuracy across various imaging modalities, including mammograms, ultrasounds, and MRIs[9,10]. Studies suggest that the transition from manual feature engineering to autonomous deep learning has revolutionized the screening process[6,7].

Multi-modal Integration and Accuracy

The findings indicate that the integration of multimodal data is essential in improving the accuracy rate of detecting cancer among patients, especially those having dense breast tissues, compared to the conventional methods used in mammography[14,15]. The AI systems, which involve the integration of different data types, prove to be more accurate, sensitive, and adaptable compared to previous systems[20,25]. The integration of DL and computer-aided detection systems improves diagnostic performance while reducing human error[18,25].

Discussion

This systematic review evaluated the effectiveness and clinical applicability of deep learning (DL)-based computer-aided diagnosis (CAD) systems for breast cancer detection. The findings indicate that convolutional neural network (CNN)-based models generally demonstrate superior diagnostic performance compared with conventional CAD systems, achieving higher accuracy, sensitivity, and specificity across multiple imaging modalities. These models are particularly effective in recognizing complex imaging patterns and detecting subtle malignant lesions, thereby supporting earlier and more accurate diagnosis[9,10].

Despite these promising findings, several challenges remain. Although DL-based systems may improve diagnostic accuracy, false-positive results may still occur, potentially leading to unnecessary follow-up examinations, biopsies, and increased patient anxiety[3,4]. Furthermore, the widespread clinical adoption of AI remains limited by concerns regarding model interpretability, transparency, and integration into existing clinical workflows[12,13].

The reviewed evidence also suggests that AI should be considered as a decision-support tool rather than a replacement for radiologists. While many radiologists have expressed positive attitudes toward AI-assisted diagnosis, concerns remain regarding workflow disruption, professional responsibilities, and the evolving role of healthcare professionals[12,13].

In addition, the integration of deep learning with CAD systems has demonstrated the potential to improve diagnostic performance through advanced image analysis and automated feature extraction[25]. However, successful implementation in clinical practice requires careful consideration of ethical and regulatory issues, including data privacy, algorithmic bias, transparency, and accountability[13,29,30]. Addressing these challenges through standardized validation, multidisciplinary collaboration, and appropriate regulatory frameworks will be essential for the safe and effective integration of AI into routine breast cancer screening and diagnosis[13,29,30].

Limitation

Although substantial advances have been achieved in AI-based diagnostics, this review has several limitations. First, the heterogeneity and lack of standardization of imaging datasets remain important challenges, as AI models developed using one population or dataset may not perform equally well in different clinical settings[9,25].

Second, the limited interpretability of deep learning models remains a major barrier to clinical implementation. Unlike conventional decision-making processes, many AI models function as "black boxes," making it difficult for clinicians to understand how diagnostic decisions are generated. This lack of transparency may reduce clinicians' trust in AI-assisted diagnosis[12,13].

Third, the high computational requirements and the need for seamless integration into existing hospital workflows continue to present practical challenges for widespread clinical adoption[12,30].

Finally, evolving ethical and regulatory issues, including algorithmic bias, patient data privacy, and accountability, require further investigation before AI-based diagnostic systems can be widely implemented in routine clinical practice[13,29,30].

Conclusion

This systematic review evaluated the role of deep learning (DL)-based computer-aided diagnosis (CAD) systems in the early detection of breast cancer. Evidence from the 13 included studies indicates that traditional CAD systems, which rely on manually extracted features, often yield higher false-positive rates, particularly in dense breast tissue[3,4,14]. In contrast, deep learning-based models, particularly CNN-based architectures, demonstrate higher diagnostic accuracy and lower false-negative rates, thereby improving early breast cancer detection[9,10,25].

These advantages are accompanied by challenges such as limited interpretability, ethical concerns, and variability in clinical acceptance[13,29,30]. Most radiologists consider AI to be a supportive tool; however, concerns remain regarding trust in AI systems and their integration into routine clinical workflows[12]. Future research should focus on the development of multimodal and hybrid models that combine the interpretability of conventional CAD systems with the flexibility and learning capability of deep learning approaches[25]. For the safe and effective clinical implementation of AI-based diagnostic systems for breast cancer, challenges related to data quality, algorithmic bias, transparency, and standardized validation must be addressed[13,29,30].

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249.

2. DeSantis CE, Ma J, Gaudet MM, et al. Breast cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(6):438-451.

3. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology*. 2019;290(2):305-314. doi: 10.1148/radiol.2018181371.

4. Dang LA, Chazard E, Poncelet E, et al. Impact of artificial intelligence in breast cancer screening with mammography. *Breast Cancer*. 2022;29(6):967-977.

5. Lång K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence (MASAI) trial: A clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *Lancet Oncol*. 2023;24(8):936-944.

6. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.

7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25:1097-1105.

8. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiology. Med Phys. 2019;46(1):e1-e36.

9. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89-94.

10. Ribli D, Horváth A, Unger Z, et al. Detecting and classifying lesions in mammograms with deep learning. *Sci Rep*. 2018;8:4165.

11. Rodríguez-Ruiz A, Lång K, Gubern-Mérida A, et al. Stand-alone artificial intelligence for breast cancer detection. *Radiology*. 2019;290(2):305-314.

12. Schaffter T, Buist DSM, Lee CI, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw Open*. 2020;3(3):e200265.

13. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*. 2019;292(1):60-66.

14. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med*. 2007;356(3):227-236.

15. Lauritzen AD, Lång K, von Euler-Chelpin M, Andersson I, Zackrisson S, Rodríguez-Ruiz A. AI-integrated screening to replace double reading of mammograms: A population-wide accuracy and feasibility study. *Radiol Artif Intell*. 2024;6(3):e230529.

16. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals. *Lancet Digit Health*. 2019;1(6):e271-e297.

17. Wang J, Yang X, Cai H, et al. Discrimination of breast cancer with CNN-based classification. *Comput Methods Programs Biomed*. 2016;134:47-55.

18. Kooi T, Litjens G, van Ginneken B, et al. Large-scale deep learning for computer-aided detection of mammographic lesions. *Med Image Anal*. 2017;35:303–312.

19. Lauritzen AD, Rodríguez-Ruiz A, von Euler-Chelpin MC, et al. An artificial intelligence-based mammography screening protocol for breast cancer: Outcome and radiologist workload. *Radiology*. 2022;304(1):41-49. doi:10.1148/radiol.210948

20. Kim HE, Kim HH, Han BK, et al. Changes in cancer detection and false-positive recall after AI support. *Radiology*. 2020;294(1):38-49.

21. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56.

22. Ahmad OF, Stoyanov D, Lovat LB. Artificial intelligence for clinical decision support. *BMJ*. 2020;368:m121.

23. Lotter W, Diab AR, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med*. 2021;27(2):244-249.

24. Lehman CD, Yala A, Schuster T, et al. Mammographic breast density assessment using deep learning. *Radiology*. 2019;290(1):38-46.

25. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.

26. Elhakim MT, Stougaard SW, Graumann O, et al. AI-integrated screening to replace double reading of mammograms: A population-wide accuracy and feasibility study. *Radiol Artif Intell.* 2024;6(6):e230529.

27. Kim HE, Kim HH, Han BK, et al. Artificial intelligence for breast cancer screening in mammography (AI-STREAM): Preliminary analysis of a prospective multicenter cohort study. *Nat Commun.* 2025;16:1897.

28. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.

29. European Commission. Ethics Guidelines for Trustworthy AI. European Commission; 2019.

30. US Food and Drug Administration. Artificial intelligence and machine learning in medical devices. FDA; 2021.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 153-162.

AI Enabled Digital Twin Approaches in Nutrition and Dietetics: Evidence, Potential, and Limitations

Burcu AKSOY CANYOLU[*], Nilüfer ŞEN[**]

Abstract

Rapid advances in technology, particularly the integration of artificial intelligence (AI) technology into our lives, have increased interest in digital twin (DT) technology, which is a dynamic virtual model of a physical system. In the healthcare sector, DT is also seen as having the potential to bring about lasting transformation in areas such as drug development, advanced diagnostics and preventive treatment, clinical research, and personalized medicine. In the coming years, DT is expected to enable personalized nutrition by integrating genetic, epigenetic, microbiome, metabolic, immunological, and lifestyle data into comprehensive virtual models. This new paradigm has the potential to provide groundbreaking opportunities for the management and prevention of various nutrition-related diseases, obesity, and support healthy aging. Although early research in the field of nutrition is promising, various challenges and limitations persist, including data standardization, privacy, data quality and security, ethical concerns, high costs, scalability, and clinical validation issues. Currently, the use of DT in the field of nutrition and dietetics remains in the proof-of-concept stage. Nevertheless, it is anticipated that as these limitations are overcome over time, this technology will transform and guide global healthcare systems. This narrative review defines the concept of DT from a healthcare perspective, summarizes its current applications in nutrition and dietetics, and outlines its high potential, key limitations, and challenges.

Keywords: Digital twin, artificial intelligence, human digital twin, personalized nutrition, data-driven nutrition science.

Beslenme ve Diyetetikte Yapay Zekâ Destekli Dijital İkiz Yaklaşımları: Kanıtlar, Potansiyel ve Sınırlılıklar

Öz

Teknolojideki hızlı ilerlemeler, özellikle yapay zeka teknolojisinin yaşamın pek çok alanına girmesiyle birlikte, fiziksel sistemlerin dinamik sanal modelleri olan dijital ikizler farklı sektörlerde kullanılmaya başlanmıştır. Hayati sektörlerden sağlık sektöründe de dijital ikiz teknolojisi; ilaç geliştirme, ileri tanı ve koruyucu tedavi, klinik araştırmalar ve kişiselleştirilmiş tıp alanlarında kalıcı dönüşüm vaat eden yüksek potansiyeller olarak görülmektedir. Gelecek yıllarda dijital ikizlerin, genetik, epigenetik, mikrobiyom, metabolik, bağışıklık ve yaşam tarzı verilerini bütüncül bir şekilde değerlendirerek kişiselleştirilmiş beslenmeyi mümkün kılması beklenmektedir. Bu yeni paradigma; obezite başta olmak üzere beslenmeyle ilişkili birçok hastalığın yönetimi, önlenmesi ve sağlıklı yaş almanın desteklenmesi açısından çığır açıcı araçlar sunma potansiyeline sahiptir. Mevcut durumda beslenme alanında yapılan çalışmalar umut vaad ederken; veri standardizasyonu, gizlilik, veri kalitesi ve güvenliği, etik kaygılar, yüksek maliyet, ölçeklenebilirlik ve klinik doğrulama gibi çeşitli zorluk ve sınırlamalar devam etmektedir. Bu nedenle beslenme ve diyetetik alanında dijital ikiz teknolojisinin kullanımı henüz kanıtlanma aşamasındadır. Ancak zaman içinde bu sınırlamaların aşılmasıyla bu teknolojinin sağlık hizmetlerini ve sistemini evrensel ölçekte dönüştürücü ve yönlendirici bir konuma geçeceği öngörülmektedir. Bu geleneksel derlemede sağlık alanı perspektifinden dijital ikiz kavramı tanımlanmış; beslenme ve diyetetik alanında dijital ikiz uygulamalarının

---

Derleme Makale (Review Article)
Geliş / Received: 23.11.2025 & **Kabul / Accepted:** 31.03.2026
DOI: https://doi.org/10.38079/igusabder.1829028
[*] Assoc. Prof., Istanbul Medeniyet University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: burcu.aksoy@medeniyet.edu.tr ORCID https://orcid.org/0000-0003-0205-4099
[**] MSc., Gazi University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Ankara, Türkiye.
E-mail: nsen397@gmail.com ORCID https://orcid.org/0000-0001-9910-0519

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

mevcut durumu sunulmuş; ayrıca yüksek potansiyelinin yanında karşılaşılan sınırlamalar ve zorluklara yer verilmiştir.

Anahtar Sözcükler: Dijital ikiz, yapay zeka, insan dijital ikizi, kişiselleştirilmiş beslenme, veri temelli beslenme.

Introduction

In 2011, a revolution emerged worldwide towards a smart Industry that combines advanced technologies, or "Industry 4.0"[1]. This transformation has affected various sectors, from agriculture to industry and healthcare[2]. Artificial intelligence (AI) has become a key driver of the digital transformation associated with Industry 4.0[3]. The rapid rise of the Internet of Things (IoT), cloud, and especially AI technology has made digital twin systems (DTs) more accessible and applicable, and DT technology has gained importance in many sectors, from academia to healthcare[4]. Today, DT, also known as digital mapping, has become a rapidly growing potential by using AI technology to collect and store data, gain insights, and create digital representations of physical objects[5]. The digital twin model has promising potential in the field of nutrition and dietetics because these models can enable personalized nutrition by combining biochemical, genetic, microbial, and lifestyle data[6,7]. Although DT models in the fields of nutrition and medicine demonstrate remarkable potential and are gaining increasing scientific attention, they are accompanied by critical limitations and challenges that must be carefully addressed[9,10].

This review aims to explain the concept of digital twins, summarize their current applications in the field of nutrition and dietetics, and evaluate their potential alongside the key limitations and challenges.

Material and Methods

This narrative review synthesizes the existing literature on the applications of DTs in the field of nutrition and dietetics. A literature review was conducted using the PubMed, Scopus, Web of Science, and Google Scholar databases, with searches performed using the keywords and combinations "digital twin," "artificial intelligence," "personalized nutrition," "nutrition," "dietetics," "microbiome," "diseases," and "healthcare." The review includes randomized controlled and observational studies published over the past 10 years (2017–2025), systematic reviews, meta-analyses, and mechanistic modeling studies that explain the mechanisms of digital twin systems and present simulation-based approaches. Animal studies, direct in vitro studies, theses, conference proceedings, and gray literature were excluded.

Concept of Digital Twin

The term 'digital twin' was first defined by Dr. Michael Grieves in an engineering presentation on Product Lifecycle Management at the University of Michigan in 2002[11]. According to this definition, digital twins are virtual representations of real systems, assets, and processes that mirror their real-world counterparts with real-time data[12,13]. The most up-to-date definition was included in the National Academies of Sciences, Engineering, and Medicine (NASEM) report, "Foundational Research Gaps and Future Directions for Digital Twins" in 2024.[14] It was described that "A DT is a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems), is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value[14]. Digital Twins have three key aspects: data collection, data modeling, and data application[15]. By using four core technologies the Internet of Things (IoT), AI, Extended Reality (XR), and Cloud Computing digital twin system can collect and store real-time data, analyze information to generate actionable insights, and create dynamic digital representations of physical entities[16]. Today, DT enables the rapid identification and resolution of physical problems across a wide range of sectors including automotive, agriculture, mining, utilities, retail, healthcare, and defense while simultaneously facilitating faster value creation, efficiency gains, and the acceleration of existing processes[4].

Digital Twin Technology in Healthcare

Digital twin technology in the healthcare sector has gained significant prominence, particularly after the COVID-19 pandemic and it seems to hold the primary transformative potential in healthcare[13]. In the broadest sense, digital twins can be identified as three-dimensional models or virtual dashboards that simulate an individual's physiology, metabolism, and biological system, including organs, hormones, and the microbiome within a dynamic virtual environment[12]. Today, DT holds great promise for transforming the healthcare sector, spanning applications in drug discovery and development, advanced diagnostics and preventive care, clinical research, personalized medicine, healthcare facility and operational design, as well as medical education and training[17,18]. In particular, AI-integrated DT can transform heterogeneous data sources such as clinical outcomes, medical imaging, and wearable sensor data into patient-specific digital models through multimodal data fusion and time-series predictive modeling[18].

Currently, DT technologies are actively being researched and applied across three primary domains: personalized medicine, operational efficiency, and scientific research[7].

Operational Efficiency: DT has the potential to reduce costs in healthcare centers and improve patients' hospital experiences. Thanks to DT, the determination of the centers' capacities and architectural planning, as well as the determination of workflow and personnel requirements, can be implemented correctly[19]. Second, DT can improve information management by integrating data from multiple points within healthcare facilities, thereby streamlining workflows and increasing operational efficiency. It can also reduce patient wait times by managing patient flow[20].

Scientific Research and Drug Development: The area of the healthcare sector where DT technologies are expected to have the greatest potential impact is clinical research methodology[17].

In particular, AI-powered DT has substantial potential for predicting treatment effectiveness, optimizing therapeutic strategies, and generating actionable insights without exposing the patient to risk[21]. Digital twins enhance patient safety by predicting the potential adverse effects of treatments and medications, while simultaneously supporting real-time clinical decision-making by identifying and recommending alternative therapeutic strategies[17].

Personalized Medicine: DT can represent real-time, dynamic models of biochemical pathways, cells, tissues, disease processes, and even the human body and its systems as a whole and it can make personalized medicine a reality[22]. Digital twins can enable the development of patient-specific therapeutic protocols by integrating thousands of variables and supporting clinical decision-making through advanced AI-driven analytics[23]. As a result, DT can model disease prognosis, evaluate potential therapeutic options, predict patient adherence to lifestyle interventions and ultimately simulate a range of possible clinical scenarios[7,19].

Digital Twin Concept and Potential in Nutrition and Dietetics

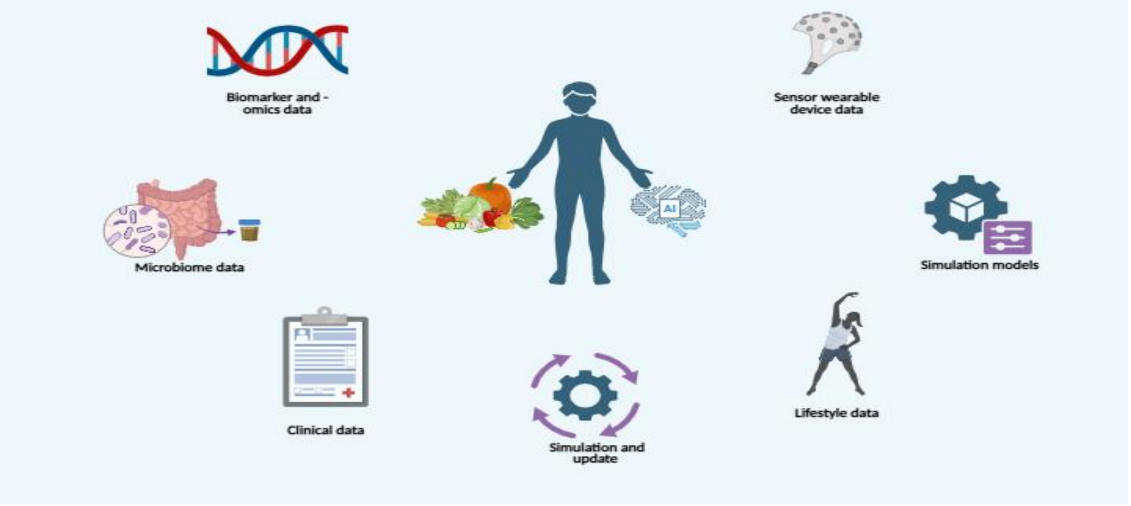
In recent years, in the field of nutrition and dietetics, AI-integrated DT models are an emerging area of research and applications that demonstrate substantial future potential[7]. These models are envisioned to monitor vulnerable populations at risk of malnutrition such as older adults, cancer patients, and individuals with nutritional deficiencies and to simulate and track patient responses in order to detect, prevent, and mitigate malnutrition risk[24]. Beyond clinical monitoring, DT technology is also expected to support telemedicine-based nutritional care[25]. In particular, DT-supported virtual consultation models may provide socio-economically or physically disadvantaged individuals, who lack access to direct nutrition counseling services, with remote and cost-effective dietary guidance[25].

It is further anticipated that future DT systems will be capable of individually predicting the nutritional value and functional effects of foods by modeling their composition and digestive behavior within the human body. In other words, a "digital twin of food" is expected to emerge in the near future[26]. In this context, the literature shows that several studies have mathematically modeled the physical and chemical transformations of food components including proteins, fibers, starches, and lipids during digestion using mechanistic equations[27,28].

These models have generated validated simulations through in vitro digestion systems, by incorporating key physiological variables such as enzyme activity, gastric emptying dynamics, and pH variations. Nevertheless, existing studies remain limited, and the concept of a fully realized "digital food twin" has yet to be achieved[26-28]. One of the areas with the greatest potential, and currently the most scientific and commercial interest in DT technology within nutrition and dietetics, is 'DT-based personalized nutrition'[8]. Thanks to DT technology, the longitudinal, high-resolution data collected through AI, including omics data, clinical and phenotypic variables, and lifestyle information, can be integrated and processed within a "human digital twin" ultimately enabling truly personalized nutrition[29]. This paradigm is expected to move far beyond generic dietary recommendations by generating insights grounded in genetic, epigenetic, microbiome-based, and real-time metabolic data, thereby offering transformative potential for obesity and a wide range of nutrition-related diseases[30].

For a human digital twin model to realize this potential in the field of nutrition, it is essential to have comprehensive biomarker data, omics profiles, microbiome information, clinical indicators, lifestyle parameters, and sensor-derived measurements[5]. Additionally, as the model evolves, it should update itself and be capable of predicting outcomes by performing scenario analyses using the entire dataset (Figure 1)[31].

Figure 1. Components of a human digital twin in nutrition and dietetics



In recent years, DT-based personalized nutrition has begun to be explored in the context of chronic and metabolic diseases and microbiome-focused research.

Type 2 Diabetes: The first investigation into the use of human DT models to improve disease prognosis in type 2 diabetes (T2DM) was conducted by Shamanna et al. In this retrospective study, a DT-enabled personalized nutrition intervention resulted in an average 24% reduction in HbA1c levels[10]. The study provided the first evidence that AI-based DT may exert meaningful effects on the clinical course of T2DM[10].

In subsequent years, Joshi et al. evaluated the effects of a DT-supported personalized intervention including individualized nutrition, physical activity, and sleep optimization on glycemic control in a randomized controlled trial involving 319 participants. After one year of follow-up, the intervention group demonstrated significantly lower HbA1c levels, with a T2DM remission rate of 72.7%[9].

A more recent large-scale real-world study (n=1985) evaluated the effects of a DT-supported personalized dietary intervention on glycemic control, medication use, and metabolic markers. After one year of follow-up, significant improvements were observed, including an average −1.8% reduction in HbA1c and a −1.5% decrease in the number of medications. In addition, the model was able to predict individuals' glycemic responses based on meal content (high-carbohydrate vs. high-protein)[32].

Silfvergren et al. developed an ordinary differential equation–based metabolic model using multiple clinical datasets, including postprandial glucose–insulin responses and fasting glucose production. The model operated across several time scales—postprandial, fasting, and long-term adaptations—and demonstrated that variations in meal frequency and fasting regimens (e.g., intermittent fasting, 5:2, small frequent meals) produce distinct effects on plasma glucose and plasma insulin[33]. Overall, these mechanistically validated models suggest future potential for predicting individual responses to meal timing, fasting patterns, and dietary composition. Such tools may eventually serve as decision-support mechanisms for dietitians, improving efficiency by reducing time and labor demands[32,33].

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Thus, although studies evaluating the impact of DT-supported personalized nutrition on disease progression and remission in T2DM remain limited, available evidence indicates that DT models can exert meaningful effects on T2DM management and that DT-enabled personalized nutrition represents a highly promising approach[30-33].

Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD): Beyond metabolic diseases, another emerging DT application area involves the gut microbiome[9]. Personalized nutrition supported by DT is a critical and increasingly prominent area of research in the management and treatment of MASLD, a disease with a rapidly increasing prevalence worldwide[9]. Joshi et al. were the first to investigate the effects of digital twin–supported personalized nutrition on MAFLD biomarkers in a randomized controlled trial (n = 319). After one year, both liver fat score and fibrosis score showed significant improvement in the DT intervention group[9].

Although several studies on DT models specific to liver diseases have been conducted, no model has yet been fully validated[34]. Existing studies remain limited, and current advancements are still at an early developmental stage[9,34].

Microbiome-based digital twin: The need to characterize an individual's microbiome in order to achieve personalized nutrition has created a clear need for microbiome-based DT models[35]. The aim of microbiome-based DT is to predict metabolic responses to individual microbiome profiles and dietary inputs, thereby enabling the optimization of nutrition and prebiotic/probiotic interventions and fully realizing personalized nutrition[35].

To the best of our knowledge, this is the only study carried out to date. Magnúsdóttir et al. conducted a study aiming to predict individuals' personalized acetate, propionate, and butyrate production potential in response to different dietary, prebiotic, and probiotic inputs. They developed "Microbial Community-Scale Metabolic Models", validated model predictions through in vitro and ex vivo experiments, and subsequently linked the SCFA profiles predicted by the model to blood-based clinical markers using large human cohort datasets[35]. This model offers the potential to predict which types of dietary fibers may be more effective and to estimate the resulting SCFA production following probiotic intake[35]. This field remains in its infancy, and further research is needed.

Digital Twin Technology in Healthcare: Challenges and Limitations

Although digital twin technology holds substantial promise in the healthcare field, it is constrained by several challenges and limitations[36].

Data standardization and quality issues, ethical concerns, difficulties in adoption by individuals, high cost, the need for large amounts of reliable and high-quality data, and data privacy and security are cited as the main limitations[13]. The lack of a legal framework for this specific issue is the main reason for concerns regarding data security, privacy, personal confidentiality, and ethical violations[24]. Furthermore, in developing countries, insufficient capacity and limitations in the healthcare system make access to DT difficult. These conditions also increase prejudice and raise ethical concerns[37]. Otherwise, current models still have limited capacity to accurately predict individualized patient responses[21,38]. Moreover, there are clear need randomized controlled trials[37]. In summary, these challenges exacerbate biases in data representation, create additional ethical concerns, and generate uncertainty regarding reliability and security among healthcare professionals and individuals[38].

Conclusion and Suggestions

DT Technology holds significant potential and transformation in the healthcare sector, especially in personalized medicine, operational efficiency and research[5]. However, there are challenges and limitations, including data integrity and security, ethical concerns, high costs, and a lack of standardization[13]. Although studies evaluating the effectiveness of models in the field of nutrition and dietetics are promising they are limited and lack randomized controlled trials. Therefore, it should be emphasized that these models are still in clinical validation phase[17]. In the future, digital twin-based approaches could serve as clinical decision support tools for dietitians. By facilitating the creation of personalized nutrition plans, they can enhance the effectiveness of patient follow-up and provide significant time and labor savings in clinical processes. These technologies can also enable the prediction of individual responses to nutrition, thereby reducing trial-and-error-based interventions and supporting the development of more targeted nutritional strategies. Moreover, thanks to real-time data integration, monitoring patients' dietary compliance can be made easier and enable early intervention when necessary. Overcoming these limitations and difficulties is possible through a multidisciplinary approach involving collaboration between healthcare professionals, policymakers and patients. In the near future, increased investment, equitable access, strengthened collaborations, standardized protocols, inclusive implementation strategies, and increased randomized controlled studies may collectively improve and eliminate the current constraints[7]. Consequently, it is believed that once the challenges associated with these high-potential DT models are overcome and adequate safeguards are established it will transform and guide the healthcare system globally.

Acknowledgements: We would like to thank BioRender for making scientific illustration easier for Figure 1.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has no financial support.

REFERENCES

1. Liao Y, Deschamps F, Loures EFR, Ramos LFP. Past, present and future of Industry 4.0: A systematic literature review and research agenda proposal. *Int J Prod Res*. 2017;55(12):3609-3629.

2. Pereira AC, Reis LP, Sousa P. Artificial intelligence and big data in nutrition and dietetics: Current applications and future perspectives. *J Acad Nutr Diet*. 2021;121(11):2267-2279.

3. Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. 4th ed. Upper Saddle River, NJ: Pearson; 2022.

4. Fuller A, Fan Z, Day C, Barlow C. Digital twin: Enabling technologies, challenges and open research. *IEEE Access*. 2020;8:108952-108971.

5. Tudor BH, Shargo R, Gray GM, et al. A scoping review of human digital twins in healthcare applications and usage patterns. *NPJ Digit Med*. 2025;8:587.

6. Mundt C, Yusufoğlu B, Kudenko D, et al. AI-Driven personalized nutrition: Integrating omics, ethics, and digital health. *Mol. Nutr. Food Res.* 2025;69.

7. Katsoulakis E, Wang Q, Wu H, et al. Digital twins for health: A scoping review. *NPJ Digit Med.* 2024;7:77.

8. Gkouskou K, Vlastos I, Karkalousos P, et al. The "virtual digital twins" concept in precision nutrition. *Adv Nutr.* 2020;11(6):1405-1413.

9. Joshi S, Shamanna P, Dharmalingam M, et al. Digital twin–enabled personalized nutrition improves metabolic dysfunction–associated fatty liver disease in type 2 diabetes: Results of a 1-year randomized controlled study. *Endocr Pract.* 2023;29(12):960-970.

10. Shamanna P, Joshi S, Shah L, et al. Type 2 diabetes reversal with digital twin technology–enabled precision nutrition and staging of reversal: A retrospective cohort study. *Clin Diabetes Endocrinol.* 2021;7(1):21. doi:10.1186/s40842-021-00134-7

11. Grieves M, Vickers J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen FJ, Flumerfelt S, Alves A, eds. *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches.* Cham, Switzerland: Springer; 2017:85-113.

12. Venkatesh KP, Brito G, Kamel Boulos MN. Health digital twins in life science and health care innovation. *Annu Rev Pharmacol Toxicol.* 2024;64:159-170.

13. Ringeval M, Etindele Sosso FA, Cousineau M, Paré G. Advancing health care with digital twins: Meta-review of applications and implementation challenges. *J Med Internet Res.* 2025;27:e69544.

14. National Academies of Sciences, Engineering, and Medicine. *Foundational Research Gaps and Future Directions for Digital Twins.* Washington, DC: National Academies Press; 2024. doi:10.17226/26894

15. Jones D, Snider C, Nassehi A, Yon J, Hicks B. Characterising the digital twin: A systematic literature review. *CIRP J Manuf Sci Technol.* 2020;29:36-52.

16. Lv Z, Xie S. Artificial intelligence in the digital twins: State of the art, challenges, and future research topics. *Digit Twin.* 2022;1(1). doi:10.12688/digitaltwin.17524.2

17. Çelik BG, Attaran M. Digital twin: Benefits, use cases, challenges, and opportunities. *Decis Anal J.* 2023;6:100165.

18. Croatti A, Gabellini M, Montagna S, Ricci A. On the integration of agents and digital twins in healthcare. *J Med Syst.* 2020;44:161-168.

19. Armeni P, Polat I, De Rossi LM, Diaferia L, Meregalli S, Gatti A. Digital twins in healthcare: Is it the beginning of a new era of evidence-based medicine? A critical review. *J Pers Med.* 2022;12(8):1255.

20. Elkefi S, Asan O. Digital twins for managing health care systems: Rapid literature review. *J Med Internet Res*. 2022;24(8):e37641.

21. Fischer RP, Volpert A, Antonino P, Ahrens TD. Digital patient twins for personalized therapeutics and pharmaceutical manufacturing. *Front Digit Health*. 2024;5:1302338.

22. Zhang K, Zhou HY, Baptista-Hon DT, et al; International Consortium of Digital Twins in Medicine. Concepts and applications of digital twins in healthcare and medicine. *Patterns*. 2024;5(8):101028.

23. Dumas M, Fay AF, Charpentier E, Matricon J. Le jumeau numérique en santé: État des lieux et perspectives d'application à l'hôpital. *Med Sci (Paris)*. 2023;39(12):953-957.

24. Abrahams M. Digital twins: The future of personalized nutrition and health? *Lifestyle Genomics*. 2025;18(1):59-63.

25. Wang Y, Lu CD, Chen W, Wang Q, Jiang H. Digital twin–enabled personalized nutrition. *Precis Nutr*. 2023;2(1):e00030.

26. Kubínová J, Přibyl J, Mach M. Personalized nutrition and digital twins of food. *Potravin Slovak J Food Sci*. 2020;14(1):1130-1136.

27. Zhu C, Bornhorst GM. Computational modeling of food digestion and bioaccessibility: Toward digital twins of food. *Food Funct*. 2021;12(8):3434-3453.

28. Liu X, Bornhorst GM. Mechanistic modeling of food digestion: Designing foods with predictable digestive behavior. *Annu Rev Food Sci Technol*. 2022;13:229-252.

29. Mathers JC. Personalised nutrition: Key concepts and current practice. *Annu Rev Nutr*. 2019;39:395-416.

30. Arshad MT, Ali MKM, Maqsood S, et al. Personalized nutrition in the era of digital health: A new frontier for managing diabetes and obesity. *Food Sci Nutr*. 2025;13(10):e71006.

31. Mosquera-Lopez C, Jacobs PG. Digital twins and artificial intelligence in metabolic disease research. *Trends Endocrinol Metab*. 2024;35(6):549-557.

32. Shamanna P, Joshi S, Thajudeen M, et al. Personalized nutrition in type 2 diabetes remission: Application of digital twin technology for predictive glycemic control. *Front Endocrinol*. 2024;15:1485464.

33. Silfvergren O, Simonsson C, Ekstedt M, Lundberg P, Gennemark P, Cedersund G. Digital twin predicting diet response before and after long-term fasting. *PLoS Comput Biol*. 2022;18(9):e1010469.

34. Zhao J, Ghallab A, Hassan R, Dooley S, Hengstler JG, Drasdo D. A liver digital twin for in silico testing of cellular and inter-cellular mechanisms in regeneration after drug-induced damage. *iScience*. 2023;27(2):108077.

35. Magnúsdóttir S, Thiele I, Heinken A, et al. Building personalized gut microbiome metabolic models to predict short-chain fatty acid production in response to diet. *Nat Microbiol.* 2022;7(9):1355-1370.

36. Vallée A. Digital twin for healthcare systems. *Front Digit Health.* 2023;5:1253050.

37. Benson M. Digital twins for predictive, preventive, personalized, and participatory treatment of immune-mediated diseases. *Arterioscler Thromb Vasc Biol.* 2023;43(3):410-416.

38. Iliuţă ME, Moisescu MA, Pop E, Ionita AD, Caramihai SI, Mitulescu TC. Digital twin: A review of the evolution from concept to technology and its analytical perspectives on applications in various fields. *Appl Sci.* 2024;14:5454.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 163-177.

An Interdisciplinary Approach to Long-Term Care: Artificial Intelligence Applications for Monitoring Nutritional Status and Social Well-Being

İrem Nur ŞAHİN ANILGAN*, Onur Zeki ANILGAN

Abstract

The global rise in the older adult population brings complex, interrelated biopsychosocial challenges, including nutritional inadequacies, sarcopenia, social isolation, and sleep disturbances. Dealing with these interconnected issues in institutional care settings is increasingly difficult given the restrictions of traditional approaches. This descriptive review analyzes, from a multidisciplinary perspective, the role of Artificial Intelligence (AI)-based technologies in monitoring nutritional status, predicting clinical risks such as malnutrition and sarcopenia, and enhancing psychosocial well-being in elder care. Relevant literature was identified through searches in PubMed/MEDLINE, Scopus, and Web of Science (2016–2026), focusing on peer-reviewed studies published in English. Current literature indicates that AI-based image-processing systems can accurately monitor dietary intake, while machine learning algorithms can enable earlier risk stratification for sarcopenia and inflammatory trajectories using biomarker data. Furthermore, social robots and non-contact sensors have been shown to reduce loneliness among older adults, improve sleep quality, and indirectly enhance motivation for eating. From a social work perspective, AI is considered an effective tool that increases organizational efficiency in case management and facilitates a shift from crisis-oriented intervention to predictive and preventive care models. The success of this technical transformation depends on establishing an ethical framework that supports privacy, dignity, and the principles of person-centered care. Overall, AI provides evidence-informed decision support for dietitians and social work professionals, helping to develop a holistic care ecosystem that optimizes the well-being of older adults.

Keywords: Artificial intelligence, elder care, nutrition, social isolation, sarcopenia, social work.

Uzun Süreli Bakımda Disiplinlerarası Bir Yaklaşım: Beslenme Durumu ve Sosyal Refahın İzlenmesinde Yapay Zekâ Uygulamaları

Öz

Yaşlı nüfusun küresel olarak artması, beslenme yetersizlikleri, sarkopeni, sosyal izolasyon ve uyku bozuklukları gibi karmaşık ve birbirleriyle ilişkili biyopsikososyal zorlukları beraberinde getirmektedir. Geleneksel yaklaşımların sınırlılıkları nedeniyle, kurumsal bakım ortamlarında bu birbiriyle bağlantılı sorunların yönetimi giderek daha güç hale gelmiştir. Bu anlatı derlemesi, yaşlı bakımında yapay zekâ (YZ) tabanlı teknolojilerin beslenme durumunun izlenmesindeki, malnütrisyon ve sarkopeni dâhil klinik risklerin öngörülmesindeki ve psikososyal refahın artırılmasındaki rolünü multidisipliner bir bakış açısıyla değerlendirmektedir. Güncel literatür, YZ destekli görüntü işleme sistemlerinin besin alımını yüksek doğrulukla izleyebildiğini; makine öğrenmesi algoritmalarının ise biyobelirteç verilerini kullanarak sarkopeni ve inflamatuvar seyre yönelik erken risk sınıflandırmasına olanak tanıdığını göstermektedir. Ayrıca sosyal robotların ve temassız sensörlerin yaşlılarda yalnızlığı azaltabildiği, uyku kalitesini

Derleme Makale (Review Article)
Geliş / Received: 28.02.2026 & **Kabul / Accepted:** 16.07.2026
DOI: https://doi.org/10.38079/igusabder.1899668
* (Corresponding Author), Res. Asst., Istanbul Beykent University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Istanbul, Türkiye. E-mail: nursahin@beykent.edu.tr ORCID https://orcid.org/0000-0002-5446-9467
** Res. Asst., Istanbul Gelişim University, Faculty of Health Sciences, Department of Social Work, Istanbul, Türkiye. E-mail: ozanilgan@gelisim.edu.tr ORCID https://orcid.org/0009-0006-4446-348X

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

iyileştirebildiği ve dolaylı olarak yeme motivasyonunu artırabildiği bildirilmektedir. Sosyal hizmet perspektifinden, YZ vaka yönetiminde kurumsal verimliliği artıran ve kriz odaklı müdahaleden öngörülü ve önleyici bakım modellerine geçişi destekleyen stratejik bir araç olarak görülmektedir. Bununla birlikte, bu teknolojik dönüşümün başarısı mahremiyeti, onuru ve kişi merkezli bakım ilkelerini koruyan bir etik çerçevenin oluşturulmasına bağlıdır. Genel olarak YZ, diyetisyenler ve sosyal hizmet uzmanları için kanıta dayalı karar desteği sunmakta ve yaşlı bireylerin iyilik hâlini optimize eden bütüncül bir bakım ekosisteminin geliştirilmesine katkı sağlamaktadır.

Anahtar Sözcükler: Yapay zeka, yaşlı bakımı, beslenme, sosyal izolasyon, sarkopeni, sosyal hizmet.

Introduction

The rapid growth of the older adult population worldwide is placing unprecedented pressure on medical care systems and institutional care models. The aging process is characterized not only by an increased prevalence of chronic diseases but also by complex, interrelated biopsychosocial problems, including nutritional inadequacies, sarcopenia, sleep disturbances, and social isolation[1]. In particular, among individuals residing in long-term care facilities, malnutrition and frailty contribute to the acceleration of inflammatory processes and a marked decline in quality of life[2]. The literature indicates that these processes are not independent; for instance, inadequate nutrition can precipitate loss of muscle mass (sarcopenia), which in turn restricts physical activity and thereby intensifies social isolation and depressive symptoms[3].

The labor-intensive nature of conventional monitoring methods, their reliance on periodic assessments, and their susceptibility to human error have made digital transformation and the embedding of Artificial Intelligence (AI) in geriatric care increasingly essential. Contemporary AI systems, through deep learning and machine learning algorithms, can detect latent patterns in clinical presentations far more rapidly and with greater accuracy than the human eye[4]. Computer vision technologies can analyze meal consumption and plate waste in real time to identify macro and micronutrient deficiencies. At the same time, wearable sensors can record sleep-wake cycles and provide early warning signals of cognitive decline[5].

Moreover, social robots and AI-based interactive systems are positioned not as substitutes for staffing shortages in nursing homes, but rather as “digital assistants” that support residents’ social well-being. In older adults with dementia, the use of therapeutic robots such as Paro has been observed to reduce agitation, improve social interaction, and even minimize the need for pharmacological interventions[6]. Situated at the intersection of nutrition/dietetics and social work, these applications serve not only as tools for clinical monitoring but also as preventive interventions that preserve dignity and optimize quality of life.

From a social work standpoint, the integration of AI into elder care can be conceptually anchored in the ecological systems approach, which frames an older adult's well-being as the product of continuous interactions across micro- (individual biological and psychological status), meso- (family, care staff, and the immediate institutional environment), and macro-level (health and social policy) systems. Within this framework, AI-derived data serve to render these interacting systems visible: individual-level indicators (e.g., nutritional intake, biomarkers, sleep patterns) can be linked to interpersonal and environmental factors (e.g., commensality, social interaction, staffing levels) to inform assessment and intervention across system levels. At the same time, deploying these technologies in line with social work values requires grounding in an empowerment- and rights-based approach, in which older adults are positioned not as passive objects of monitoring but as active agents who retain autonomy, dignity, and control over their own data and care decisions. Situating AI applications within these theoretical and ethical frameworks ensures that technological efficiency remains subordinate to the profession's core commitment to person-centered care and social justice[7,8].

Accordingly, this review assesses, from a multidisciplinary perspective and based on current evidence, the role of AI-based systems in long-term care settings for monitoring nutritional status, predicting clinical risks such as sarcopenia and malnutrition, and safeguarding social well-being.

Material and Methods

This study was designed as a narrative (descriptive) review aiming to synthesize current evidence on the role of AI-based technologies in monitoring nutritional status and supporting psychosocial well-being in long-term care. Literature searches were conducted in the PubMed/MEDLINE, Scopus, and Web of Science databases, covering publications from January 2016 to February 2026. The search combined keywords related to three domains: (i) population and setting ("older adults", "long-term care", "nursing home", "elder care"), (ii) technology ("artificial intelligence", "machine learning", "deep learning", "social robot", "smart sensor"), and (iii) outcomes ("nutritional status", "malnutrition", "sarcopenia", "social isolation", "sleep", "psychosocial well-being"). Priority was given to peer-reviewed original studies, systematic reviews, and meta-analyses published in English. Studies were included based on their relevance to the interdisciplinary intersection of nutrition/dietetics and social work; theoretical, technical, and conceptual sources that contextualized the clinical findings were also incorporated. Given the descriptive nature of the review, no formal quality-appraisal or risk-of-bias assessment was performed.

Clinical Monitoring and Artificial Intelligence in Older Adults

One of the most important components of clinical management in long-term care facilities is the prompt recognition of nutritional inadequacies (malnutrition) and the subsequent development of sarcopenia. Conventional screening instruments (e.g., MNA, NRS-2002) typically provide static data and often yield actionable findings only once the condition has become established or chronic. However, recent studies indicate that AI-based predictive systems can detect risk at much earlier stages, thereby permitting more timely and preemptive intervention[4].

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Digital Monitoring of Dietary Intake and Image Processing

In older adults, monitoring meal consumption is typically based on staff-recorded "plate waste" observations, a process that is largely manual and highly subjective. Recent AI-based computer vision technologies, however, have substantially digitized this workflow, thereby minimizing errors. Deep learning structures, particularly convolutional neural networks (CNNs), can analyze pre- and post-meal plate photographs within seconds to identify food type and portion volume and, consequently, estimate energy intake and macronutrient consumption, with reported accuracy rates reaching up to 94% in older adults in long-term care and hospital settings[9–11]. Notably, systems such as MealVision have been shown to reduce dietitians' manual data-entry burden by approximately 80% while allowing real-time reporting of inadequate protein intake, a key trigger for sarcopenia[12,13].

AI applications are not limited to visual analytics; they also track dietary intake through smart utensils and wearable sensors. Smart forks and spoons can monitor eating speed and bite count, thereby enabling the assessment of physical risks, such as rapid eating or chewing/swallowing difficulties (dysphagia)[14]. In addition, micro-electromagnetic sensors placed on the neck or acoustic (sound-based) recognition systems are able to analyze swallowing sounds to inform social work professionals about an individual's daily fluid intake or the frequency of skipped meals[15,16].

To protect privacy in residential care settings, "ambient intelligence" systems are increasingly preferred over camera-based monitoring[17]. IoT (Internet of Things) sensors embedded in kitchen cabinets or refrigerators, when combined with AI algorithms, can track changes in kitchen-related activities (e.g., the refrigerator being opened less frequently than usual) and flag possible risks such as concealed anorexia or cognitive decline[18]. These data can help social work professionals better understand the correlation between loneliness and reduced appetite, hence supporting timely psychosocial interventions.

Assessment of Sarcopenia and Inflammatory Processes

Sarcopenia is strongly associated with adverse outcomes in older adults, including a higher risk of falls and fractures, as well as downstream functional decline that can compromise independent living. Conventional diagnostic approaches such as dual-energy X-ray absorptiometry (DXA) and bioelectrical impedance analysis (BIA) typically require clinic-based assessment and provide only snapshot measurements that do not reflect short-term changes in inflammatory activity. In contrast, machine-learning approaches can integrate routinely collected blood biomarkers, among them C-reactive protein (CRP) and pro-inflammatory cytokines such as interleukin-6 (IL-6) and tumor necrosis factor-alpha (TNF-α), with clinical and functional variables to support earlier risk stratification and prediction of sarcopenia-related trajectories, assisting timely nutritional intervention targeted at mitigating inflammation-driven muscle catabolism[19,20].

AI-based ultrasound analysis is also transformative for the diagnosis of sarcopenia. Whereas manual ultrasound measurements are highly susceptible to operator-related variability, AI algorithms can automatically evaluate muscle echogenicity (an indicator of muscle quality) and architectural characteristics (e.g., pennation angle), thereby producing quantitative reports of muscle loss[21]. One study also emphasizes that AI can integrate these radiologic features with dietary data, particularly leucine and vitamin D intake, to estimate an individual's "muscle regeneration capacity"[22].

In parallel, the age-related process of chronic, low-grade inflammation (inflammaging) can be examined using AI-based data-mining approaches that jointly evaluate clinical findings, biomarkers (e.g., CRP, IL-6), dietary records, and genetic variants. This integrated analytic framework can help identify which genomic profiles are most likely to show improvements in inflammatory markers and derive greater clinical benefit from Mediterranean-style eating patterns or other anti-inflammatory dietary strategies[23]. Those insights support the development of care plans for older adults in social care settings that are not simply "adequate," but optimized and explicitly anti-inflammatory.

Psychosocial Well-Being and Technological Support

In later life, the preservation of psychosocial well-being is as critical as physical health parameters for sustaining general well-being and stabilizing nutritional status. Among older adults receiving institutional care, social isolation and the chronic sense of loneliness that often accompanies it not only threaten mental health but also cause physiological consequences, including diminished appetite, skipped meals, and sleep disturbances[24]. In this context, AI-based applications deliver comprehensive support by functioning both as passive monitoring mechanisms and as active intervention tools.

The Negative Cycle Between Social Isolation and Nutrition

In later life, social isolation is not simply a psychological condition; it also constitutes a biological risk factor that can precipitate nutritional inadequacy and physical decline. The literature links "eating alone" (i.e., reduced commensality) directly to lower overall food intake and poorer meal quality among older adults[25]. Within institutional care places where social interaction is limited, individuals' motivation to eat may diminish, thereby initiating a cycle of appetite loss commonly described as the "anorexia of aging," a term referring to the physiological and psychosocial decline in appetite and food intake that frequently accompanies advancing age[26].

AI-based analytic systems can detect this negative cycle before it crystallizes into a clear clinical presentation. AI-based emotion recognition and voice-analysis technologies can monitor speech rate, lexical diversity, and melancholic shifts in vocal tone to quantify levels of social withdrawal[27]. For example, an older adult's preference for staying in their room rather than attending the dining hall, or abrupt decreases in the duration of social interactions, may be labeled by AI algorithms as "risk of social isolation". At the same time, concurrent effects on dietary intake (e.g., skipped meals, reduced portion size) are reported[28].

In addition, interaction data collected via therapeutic robots and digital companions can provide social work professionals with quantitative indicators of an individual's level of "social hunger". Through studying the quality of an older person's bond with a robot or digital assistant, AI can identify the times of day when loneliness peaks and determine how these periods correlate with eating behaviors[29]. This evidence-based approach enables social work professionals to design targeted "social support interventions" tailored not only to general activities but also to meal times when individuals may be most vulnerable, thereby providing preventive protection against loneliness-related nutritional disturbances[30].

Monitoring Sleep Quality with Smart Sensors and Its Relationship with Nutrition

Sleep disturbances in older adults can disrupt circadian rhythms and adversely affect hormonal balance that regulates appetite, particularly ghrelin & leptin levels. Poor-quality or insufficient sleep may reduce metabolic rate and, the following day, contribute to blunted appetite that can culminate in the "anorexia of aging" or lead to irregular and inferior food choices[31]. Conventional sleep-monitoring methods such as polysomnography are invasive, could aggravate agitation in older individuals, and are difficult to implement in institutional care settings[32].

In this context, AI-based non-contact sensors and smart-bed technologies offer a disruptive and minimally burdensome approach to monitoring sleep hygiene in older adults[33]. AI algorithms can continuously analyze sleep phases, respiratory rate, heart rate variability (HRV), and involuntary nocturnal movements to generate an integrated "sleep score"[34]. For instance, when AI detects REM-sleep fragmentation or repeated awakenings, these data can be linked to nutritional monitoring interfaces. If a sustained decline in sleep quality is observed, the system can send early alerts to care staff, indicating that the pattern may increase malnutrition risk via appetite loss[32].

Moreover, AI models integrated with smart-home systems and wearable technologies can go beyond monitoring sleep quality by optimizing the sleep surroundings (e.g., light levels and room temperature) in accordance with an individual's biological clock, therewith supporting circadian regulation[35]. From a social work perspective, this electronic intervention represents a non-pharmacological strategy to improve sleep quality, enabling individuals to start their morning meals feeling more rested and with greater appetite. In turn, it can act as a critical defensive shield against the development of frailty syndrome associated with undernutrition[36].

Digital Companions and Enhancing Social Interaction

Therapeutic robots and voice-based AI assistants are increasingly integrated into institutional care as "digital companions" to reduce the harmful effects of social isolation among older adults[37,38]. Particularly in individuals living with dementia or cognitive decline, the use of biomimetic robots such as *Paro* has been associated with significant reductions in agitation and anxiety, therewith enhancing residents' capacity to engage with their surroundings[6,39]. This technology-based intervention not only provides psychological relief but may also reduce disruptive behaviors noted during meals (e.g., wandering, refusal to eat), thus rendering food intake more calm and efficient[6,39].

AI-based interactive systems can further provide cognitive and social stimulation through verbal and/or embodied interaction, and many are designed to support daily self-care via personalized prompts (e.g., hydration, medicine schedules, and meal-time reminders)[38,39]. In addition, AI algorithms can monitor reaction time and lexical richness during these communications and deliver real-time reports to social work professionals regarding possible declines in cognitive functioning[40].

From a social work perspective, digital companions are not intended to replace staff workload; rather, they provide older adults with a sense of “companionship” and “safety” during periods when personnel are not physically present[41]. This form of “continuous social presence” can help disrupt the appetite-loss cycle triggered by loneliness, strengthen older adults’ adaptation to institutional living, and make an indirect yet substantial contribution to the sustainability of nutritional status[2]. Ultimately, these technologies are positioned as strategic adjunct tools in elder care, supporting rather than supplanting human touch while improving biopsychosocial well-being[5].

AI Applications from a Social Work Perspective

The discipline of social work seeks to protect individuals’ biopsychosocial integrity and improve social welfare. In long-term care settings, the deployment of artificial intelligence provides social work professionals with analytics-based capacity in case management, risk assessment, and resource allocation[42]. Whereas intervention processes in traditional care models are largely reactive (i.e., initiated after an event occurs), AI enables a more proactive orientation through facilitating earlier detection and preventive action[43].

Organizational Effectiveness and Clinical Decision Support Systems

Institutional elder care is an operational context with low tolerance for error, requiring the management of complex health needs under limited staffing resources. In conventional care models, a substantial proportion of staff time is devoted to routine data entry and documentation, which in turn limits opportunities for direct social interaction with residents[44]. AI-based Clinical Decision Support Systems (CDSS) can improve these administrative demands, thereby improving organizational efficiency and also boosting the overall quality of care[45].

AI algorithms can concurrently analyze electronic health records (EHRs), nutrition-tracking interfaces, and sensor-derived data to generate “smart task lists” for social work professionals and dietitians[46]. For instance, the system may integrate indicators such as sudden weight loss, deteriorating sleep quality, and decreased interaction with a social robot, and subsequently issue a “high-priority case” alert to relevant professionals[47]. Rather than citing a single fixed percentage improvement, the literature more consistently supports that well-designed dashboards and digital decision-support can improve task execution time, decision-making, and conformance to evidence-based processes, depending on usability and implementation quality[48].

Moreover, the objective datasets generated by AI-based systems can shift multidisciplinary in-house meetings (case conferences) from personal interpretations toward evidence-informed analyses. Social workers, dietitians, and physicians can jointly review an older adult's status via a shared, digital, and up-to-date dashboard, permitting timely revisions to intervention plans[48]. In this sense, the digital ecosystem functions not only as organizational automation but also as a safety net that helps guarantee continuity and accuracy in the services provided to older adults[45].

Predictive Social Work and Preventive Intervention

Traditional social work practice is frequently reactive in nature, typically being activated only after a problem has emerged (e.g., falls, acute malnutrition, severe depression). In contrast, machine learning-based predictive analytics models can identify low-intensity alterations within big datasets, thereby creating a "pre-crisis" window for social work[48]. This approach constitutes a new paradigm in the discipline, often referred to as predictive social work, which aims to manage risks before they materialize[4].

AI algorithms can integrate subtle reductions in appetite, micro-changes in sleep duration, and patterns of social withdrawal into a single composite risk score. For example, a slight slowing of gait speed coupled with decreased protein intake may be flagged by the system as an "impending frailty episode" or "elevated fall risk"[46,47]. From a social work perspective, this shifts the focus from post-event rehabilitation following injury or illness to the prompt initiation of preventive interventions such as modifying the individual's living environment or increasing nutritional support without delay[42].

Predictive systems can likewise play a key role in preventing adverse mental health outcomes such as severe depression or suicide by monitoring the impact of loneliness on older adults through digital interaction data[28]. These AI-based "early warnings" enable social work professionals to move beyond intuitive decision-making toward evidence-informed, appropriately timed interventions, which have been associated with meaningful reductions in hospitalization rates and morbidity in nursing home settings[42]. Within this integrated framework, the social worker's role shifts from that of a "crisis responder" to a proactive "architect of well-being," responsible for actively shaping and sustaining older adults' quality of life[2].

Technology, Social Justice, and Access to Services

Artificial intelligence is increasingly positioned as a strategic tool for overcoming the physical, geographic, and socioeconomic barriers that older adults face in accessing health and social services[43]. In line with the core social work principle of social justice, AI can "democratize" access to high-quality monitoring and intervention services for disadvantaged groups, older adults with disabilities, and long-term care facilities in rural or under-resourced settings where staffing is limited[42]. Remote monitoring technologies and AI-based telehealth systems allow specialists to extend care beyond institutional boundaries and reach broader populations in accordance with the principle of "aging in place"[2,43].

The objective and continuous datasets generated by AI algorithms may also help minimize implicit bias and discriminatory practices in service delivery, thereby supporting the provision of fair and equitable care plans tailored to each older adult's level of need[5,41]. In traditional systems, more "active" or outspoken residents may receive disproportionate attention; by contrast, AI-based systems can identify even the most "invisible" individuals who silently experience social isolation or appetite loss, bringing them to the attention of social work professionals[6]. This enables the rational allocation of limited social work resources based on need, thereby strengthening institutional equity.

In addition, AI-based natural language processing and translation technologies can help older adults from diverse ethnic backgrounds or those experiencing language barriers accurately communicate their dietary preferences and social needs to care staff[43]. From a social work perspective, such technological integration can function as an advocacy tool to reduce the digital divide, supporting older adults in becoming not merely passive recipients of care but active agents with voice and authority over their health data and rights[2]. Overall, AI serves as a lever that promotes health equity, universalizes access to services, and aligns the humanistic values of social work with the opportunities of the digital era[3].

Ethical Debates and Challenges

The deployment of artificial intelligence and digital monitoring systems into long-term care facilities introduces profound socio-ethical dilemmas that extend beyond the operational benefits these technologies may offer. The pace of technological innovation often outstrips the development of legal and ethical structures designed to regulate its impact on human values[43]. Accordingly, the use of AI in institutional care necessitates scrutiny across core ethical domains, including privacy protection, individual autonomy, and algorithmic impartiality.

Image-processing technologies and biometric sensors used for real-time monitoring of nutritional status and sleep quality require a delicate balance between an older adult's right to privacy and the pursuit of improved care quality[42]. Continuous digital surveillance may risk exacerbating "institutionalization trauma" and undermining individuals' sense of control over their own living space[6]. To address these concerns, the literature recommends prioritizing non-camera-based solutions and privacy-preserving modalities such as anonymized data streams (e.g., silhouette-based analysis or thermal sensors) and restricting data access to authorized professionals only[5].

The clinical success of such systems largely depends on the attitudes of older adults and care staff toward technology. "Technophobia" and disparities in digital literacy may lead some older individuals to resist AI-based robots or wearable biosensors. From a social work perspective, it is essential that technology does not become a coercive instrument and that individuals' right to informed consent is safeguarded[3]. In addition, if care staff perceive AI not as a supportive tool but as a threat to their professional roles, this can negatively affect system effectiveness and data entry quality[42].

AI algorithms also carry the risk of reproducing sociocultural biases embedded in the datasets on which they are trained. This may result in inaccurate risk assessments for older adults from diverse dietary cultures[4]. A central ethical concern is the possibility that digitalization could weaken the "human touch" and empathic bonds that lie at the heart of social work practice. Consequently, it is ethically imperative to position AI as an auxiliary mechanism for data generation, while ensuring that final clinical decisions and emotional-support processes remain under the oversight of human professionals (e.g., social workers and dietitians) through a human-in-the-loop approach[43].

Conclusion and Recommendations

The integration of artificial intelligence into long-term care facilities demonstrates transformative potential, encompassing applications from nutritional status monitoring to the protection of psychosocial well-being. The literature reviewed indicates that AI-based systems can identify clinical risks, such as malnutrition and sarcopenia, at early stages and support interventions for complex issues, including loneliness and sleep disturbances, through social robots and smart sensors. However, the sustainability of these efficiency gains depends on maintaining a balance with privacy, ethical standards, and the principle of person-centered care.

In summary, AI serves as an essential decision-support mechanism for social work professionals and dietitians in addressing interconnected challenges such as the anorexia of aging, social isolation, and cognitive decline. In the future, AI-based smart nursing homes may evolve into holistic living environments that actively protect dignity, autonomy, and social connectedness through technology. In the Turkish context, these considerations carry particular relevance. Institutional elder care in Türkiye is delivered largely through nursing homes and elderly care and rehabilitation centers operating under the Ministry of Family and Social Services, alongside a growing number of private facilities. As the country undergoes rapid demographic aging, these institutions face increasing pressure from workforce shortages and rising care needs precisely the conditions under which AI-based decision-support and monitoring tools could offer the greatest value. Integrating such systems into the operational workflows of these centers, in a manner consistent with national data-protection legislation and the person-centered principles of the Ministry's elder-welfare policies, could support earlier detection of malnutrition and frailty, more equitable allocation of limited staff resources, and improved continuity of care. Realizing this potential will require locally validated tools, adequate digital infrastructure, and staff training tailored to the Turkish care setting. The success of this transformation will depend on integrating data-driven approaches with the empathetic values central to social work practice.

Limitations

Although the reviewed literature highlights considerable potential, several limitations temper an uncritical adoption of AI in long-term care. First, the empirical evidence is not uniformly positive: trials of socially assistive and therapeutic robots have yielded mixed or null results, with some studies reporting no significant improvement in loneliness, agitation, or quality of life, and effects that diminish once the novelty period passes or that depend heavily on the presence of a human facilitator. Second, algorithmic limitations remain substantial; models trained on non-representative datasets can produce biased or inaccurate risk estimates for underrepresented groups, and their real-world performance often falls short of the accuracy reported under controlled conditions. Concerns regarding the transparency of "black-box" models, data-privacy risks, and the potential erosion of human contact further constrain implementation. Finally, as a narrative review, this study did not employ a systematic search protocol or formal quality appraisal, which may introduce selection bias and limit the generalizability of its conclusions. These considerations underscore that AI should be regarded as a complement to, rather than a substitute for, human-centered professional judgment.

REFERENCES

1. Maleki Varnosfaderani S, Forouzanfar M. The Role of AI in hospitals and clinics: transforming healthcare in the 21st Century. *Bioengineering*. 2024;11(4):337.

2. Kurra RK, Molli SM. AI in nutrition, dental, and healthcare: transforming dietary science for comprehensive well-being. In: Modernizing the Food Industry: AI-Powered Infrastructure, Security, and Supply Chain Innovation. IGI Global Scientific Publishing; 2026:305-328.

3. Janssen SMW, Bouzembrak Y, Tekinerdogan B. Artificial intelligence in malnutrition: a systematic literature review. *Advances in Nutrition*. 2024;15(9):100264.

4. Bernstein AM, Janeke P, Riggs R V, et al. Artificial intelligence-based hospital malnutrition screening: validation of a novel machine learning model. *Clinical Informatics*. 2025;16(5):1646-1657. doi: 10.1055/a-2635-3158

5. Støre SJ, Beckman L, Jakobsson N. The effect of robot interventions on sleep in adults: a systematic review and network meta-analysis. *Journal of Clinical Sleep Medicine*. 2022;18(7):1877-1884. doi:10.5664/jcsm.10022

6. Rashid N, Leow Y, Klainin-Yobas P, Itoh S, Wu VX. The effectiveness of a therapeutic robot “Paro”, on behavioural and psychological symptoms, medication use, total sleep time and sociability in older adults. *International Journal of Nursing Studies*. 2023;145:104530.

7. Pfisterer K, Amelard R, Boger J, Keller H, Chung A, Wong A. Enhancing food intake tracking in long-term care with automated food imaging and nutrient intake tracking (AFINI-T) technology: validation and feasibility assessment. *JMIR Aging*. 2022;5(4):e37590. doi: 10.2196/37590

8. Kim JH, Lee DS, Kwon SK. Food classification and meal ıntake amount estimation through deep learning. *Applied Sciences* 2023;13(9):5742. doi: 10.3390/app13095742

9. Papathanail I, Brühlmann J, Vasiloglou MF, et al. Evaluation of a novel artificial intelligence system to monitor and assess energy and macronutrient ıntake in hospitalised older patients. *Nutrients*. 2021;13(12):4539. doi: 10.3390/nu13124539

10. Amini N, Devriendt A, Lapauw L, et al. Estimating protein intake in sarcopenic older adults: combining food diaries and weighed powders versus 24-hour urine collections. *Journal of Nutrition, Health and Aging*. 2025;29(3). doi: 10.1016/j.jnha.2024.100474

11. The Ingenious AI That's Combatting Malnutrition And Optimising Nutrition in Australia | Deloitte Australia. Accessed February 28, 2026. https://www.deloitte.com/au/en/services/consulting/case-studies/ingenious-ai-combatting-malnutrition-optimising-nutrition-australia.html

12. Hermsen S, Mars M, Higgs S, Frost JH, Hermans RCJ. Effects of eating with an augmented fork with vibrotactile feedback on eating rate and body weight: a randomized controlled trial. *Int J Behav Nutr Phys Act*. 2019;16(1):90. doi: 10.1186/s12966-019-0857-7

13. Bi Y, Lv M, Song C, Xu W, Guan N, Yi W. AutoDietary: A Wearable Acoustic Sensor System for Food Intake Recognition in Daily Life. *IEEE Sens J*. 2016;16(3).

14. Park T, Mahmud TI, Lee J, et al. A machine-learning-enabled smart neckband for monitoring dietary intake. *PNAS Nexus*. 2024;3(5). doi: 10.1093/pnasnexus/pgae156

15. Jovanovic M, Mitrov G, Zdravevski E, et al. Ambient assisted living: scoping review of artificial ıntelligence models, domains, technology, and concerns. *J Med Internet Res*. 2022;24(11):e36553. doi: 10.2196/36553

16. Grammatikopoulou M, Lazarou I, Alepopoulos V, et al. Assessing the cognitive decline of people in the spectrum of AD by monitoring their activities of daily living in an IoT-enabled smart home environment: a cross-sectional pilot study. *Front Aging Neurosci*. 2024;16:1375131. doi: 10.3389/fnagi.2024.1375131

17. Yeung SSY, Reijnierse EM, Pham VK, et al. Sarcopenia and its association with falls and fractures in older adults: A systematic review and meta-analysis. *J Cachexia Sarcopenia Muscle*. 2019;10(3):485. doi: 10.1002/jcsm.12411

18. Ceyhan AB, Altay O, Zhang C, Temel SG, Turkez H, Mardinoglu A. Unravelling the complexity of sarcopenia through a systems biology approach. *International Journal of Molecular Science*. 2025;26(17):8527. doi: 10.3390/ijms26178527

19. Yi J, Shin YR, Hahn S, Lee YH. Deep learning based sarcopenia prediction from shear-wave ultrasonographic elastography and gray scale ultrasonography of rectus femoris muscle. *Scientific Reports*. 2022;12(1):3596-. doi: 10.1038/s41598-022-07683-6

20. Cormier H, Rudkowska I, Lemieux S, Couture P, Voh MC. Expression and sequence variants of inflammatory genes; effects on plasma inflammation biomarkers following a 6-week supplementation with fish oil. *International Journal of Molecular Sciences*. 2016;17(3):375. doi: 10.3390/ijms17030375

21. Calcaterra V, Verduci E, Milanta C, et al. The benefits of the Mediterranean diet on ınflamm-aging in childhood obesity. *Nutrients*. 2024;16(9):1286. doi: 10.3390/nu16091286

22. de Jong Gierveld J, van Tilburg T. The De Jong Gierveld short scales for emotional and social loneliness: Tested on data from 7 countries in the UN generations and gender surveys. *Eur J Ageing*. 2010;7(2):121-130. doi: 10.1007/s10433-010-0144-6

23. Yamada Y, Shinkawa K, Nemoto M, Arai T. Automatic assessment of loneliness in older adults using speech analysis on responses to daily life questions. *Front Psychiatry*. 2021;12:712251. doi: 10.3389/fpsyt.2021.712251

24. Björnwall A, Colombo PE, Sydner YM, Neuman N. The impact of eating alone on food intake and everyday eating routines: A cross-sectional study of community-living 70- to 75-year-olds in Sweden. *BMC Public Health*. 2024;24(1):2214.

25. Wyman C, Thomas J, Lawless M, Yaxley A. Associations between nutritional and physical outcomes of community-dwelling older adults eating alone, versus with others: A systematic review. *Appetite*. 2026;217. doi: 10.1016/j.appet.2025.108327

26. Yoo IJ, Park DH, Lee OE, Park A. Investigating older adults' use of a socially assistive robot via time series clustering and user profiling: descriptive analysis study. *JMIR Form Res*. 2024;8(1):e41093. doi: 10.2196/41093

27. Yang Y, Wang C, Xiang X, An R. AI applications to reduce loneliness among older adults: a systematic review of effectiveness and technologies. *Healthcare*. 2025;13(5):446.

28. Lin J, Jiang Y, Wang G, et al. Associations of short sleep duration with appetite-regulating hormones and adipokines: A systematic review and meta-analysis. *Obesity Reviews*. 2020;21(11):e13051. doi: 10.1111/obr.13051

29. Dichter MN, Dörner J, Wilfling D, et al. Intervention for sleep problems in nursing home residents with dementia: A cluster-randomized study. *Int Psychogeriatr*. 2024;36(10):965-978. doi: 10.1017/S1041610223004489

30. Tran VP, Al-Jumaily AA, Islam SMS. Doppler radar-based non-contact health monitoring for obstructive sleep apnea diagnosis: a comprehensive review. *Big Data and Cognitive Computing*. 2019;3(1):3. doi: 10.3390/bdcc3010003

31. Siyahani F, Molina GG, Barr S, Mushtaq F. Performance evaluation of a smart bed technology against polysomnography. *Sensors (Basel)*. 2022;22(7):2605.

32. Giggins OM, Doyle J, Hogan K, George M. The impact of a cycled lighting intervention on nursing home residents: a pilot study. *Gerontol Geriatr Med*. 2019;5:2333721419897453.

33. Baandrup L, Jennum PJ. Effect of a dynamic lighting intervention on circadian rest-activity disturbances in cognitively impaired, older adults living in a nursing home: A proof-of-concept study. *Neurobiol Sleep Circadian Rhythms*. 2021;11(3):100067.

34. Marziali RA, Franceschetti C, Dinculescu A, et al. Reducing loneliness and social isolation of older adults through voice assistants: literature review and bibliometric analysis. *J Med Internet Res*. 2024;26(1):e50534. doi: 10.2196/50534

35. Fan W, Zhao R, Liu X, Ge L. Intelligent robot interventions for people with dementia: systematic review and meta-analysis of randomized controlled trials. *J Med Internet Res*. 2025;27:e59892. doi: 10.2196/59892

36. Broadbent E, Loveys K, Ilan G, et al. ElliQ, an AI-based social robot to alleviate loneliness: progress and lessons learned. *JAR life*. 2024;13:22. doi: 10.14283/jarlife.2024.2

37. Igarashi T, Iijima K, Nitta K, Chen Y. Detailed analysis of responses from older adults through natural speech: comparison of questions by AI agents and humans. *Int J Environ Res Public Health*. 2024;21(9):1170. doi: 10.3390/ijerph21091170

38. Barrett MA, Allen A, Vuong VT, et al. Impact of an artificial intelligence and machine learning enhanced electronic health record system on quality measures in nursing homes: a difference-in-differences analysis. *Journal of the American Medical Directors Association*. 2025;26(7):105680. doi: 10.1016/j.jamda.2025.105680

39. Sapci AH, Sapcı HA. Innovative assisted living tools, remote monitoring technologies, artificial intelligence-driven solutions, and robotic systems for aging societies: systematic review. *JMIR Aging,* 2019;2(2):e15429. doi: 10.2196/15429.

40. Gesner E, Dykes PC, Zhang L, Gazarian P. Documentation burden in nursing and its role in clinician burnout syndrome. *Appl Clin Inform*. 2022;13(5):983.

41. Lapp L, Egan K, McCann L, Mackenzie M, Wales A, Maguire R. Decision support tools in adult long-term care facilities: scoping review. *J Med Internet Res*. 2022;24(9):e39681.

42. Ye J, Woods D, Jordan N, Starren J. The role of artificial intelligence for the application of integrating electronic health records and patient-generated data in clinical decision support. *AMIA Summits on Translational Science Proceedings*. 2024;2024:459.

43. Abdellatif A, Bouaud J, Lafuente-Lafuente C, Belmin J, Séroussi B. Computerized decision support systems for nursing homes: a scoping review. *J Am Med Dir Assoc*. 2021;22(5):984-94. doi: 10.1016/j.jamda.2021.01.080

44. Bucalon B, Shaw T, Brown K, Kay J. State-of-the-art dashboards on clinical ındicator data to support reflection on practice: scoping review. *JMIR Med Inform*. 2022;10(2):e32695.

45. Kumar D, Suthar N. Predictive analytics and early intervention in healthcare social work: a scoping review. *Soc Work Health Care*. 2024;63(4-5):208-229.

46. Zheng X, Zeng Z, S van Schooten K, Yang Y. Machine learning approach for frailty detection in long-term care using accelerometer-measured gait and daily physical activity: model development and validation study. *JMIR Aging*. 2025;8(1):e77140.

47. Ruiz-Ruiz L, Jimenez AR, Garcia-Villamil G, Seco F. Detecting fall risk and frailty in elders with inertial motion sensors: a survey of significant gait parameters. *Sensors*. 2021;21(20):6918. doi: 10.3390/s21206918

48. González-Castro A, Benítez-Andrades JA, González-González R, Prada-García C, Leirós-Rodríguez R. Predicting fall risk in older adults: A machine learning comparison of accelerometric and non-accelerometric factors. *Digit Health*. 2025;11:20552076251331750.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 178-186.

Artificial Intelligence and Nursing in Management of Diabetes Mellitus

Gamze YEĞİN*, Songül KARADAĞ**

Abstract

The morbidity, mortality, and economic burden associated with diabetes mellitus and its complications make it an important health concern. Artificial intelligence, becoming prominent in today's advancing technology, refers to application systems or machines capable of imitating human cognitive functions and enhancing their performance with the data they acquire. Artificial intelligence can provide data-oriented, sensitive treatment and care methods in addition to existing methods, resulting in a paradigm shift in management of diabetes mellitus. Artificial intelligence-based virtual assistants offer people 24/7 service and answer their questions instantly, send reminders, and provide educational materials. They play an important role particularly in monitoring patients' blood glucose levels and reminding them to take their medications. Continuous and individualized support helps patients comply with treatment and keep their blood glucose levels under control; thus, enhancing their quality of life and ability to cope with diabetes mellitus.

Keywords: Diabetes management, artificial intelligence, nursing.

Diyabet Yönetiminde Yapay Zeka ve Hemşirelik

Öz

Diyabet ve diyabetik komplikasyonların yol açtığı morbidite, mortalite ve bunlara bağlı gelişen ekonomik yük, diyabetin önemli bir sağlık sorunu haline gelmesine neden olmaktadır. Günümüz gelişen teknolojisinde öne çıkan yapay zeka, insanın bilişsel işlevlerini taklit edebilen, elde ettiği veriler ile kendini geliştirme niteliklerine sahip uygulama sistemleri veya makineler anlamına gelmektedir. Yapay zekâ, diyabet bakımında bir paradigma değişimine yol açarak mevcut yöntemlere ek olarak veri odaklı hassas tedavi ve bakım yöntemleri sağlayabilmektedir. Yapay zeka destekli oluşturulan sanal asistanlar 7/24 erişilebilir olup, hastaların anlık sorularına yanıt verebilmekte, hatırlatıcılar gönderebilmekte ve eğitim materyalleri sunabilmektedir. Özellikle hastaların kan şekeri seviyelerini izlemede ve ilaç alımını hatırlatmada önemli rol oynamaktadır. Sürekli ve bireyselleştirilmiş destek sunulması, hastaların tedaviye uyumunu artırmakta ve kan şekeri kontrolünü iyileştirmektedir. Bunun sonucunda hastaların yaşam kalitesi artmakta ve diyabetle başa çıkma yetenekleri gelişmektedir.

Anahtar Sözcükler: Diyabet yönetimi, yapay zeka, hemşirelik.

Derleme Makale (Review Article)
Geliş / Received: 28.01.2026 & **Kabul / Accepted:** 29.06.2026
DOI: https://doi.org/10.38079/igusabder.1873892
* PhD Student, Çukurova University, Institute of Health Sciences, Department of Nursing, Adana, Türkiye.
E-mail: gamzeyegin@windowslive.com ORCID https://orcid.org/0009-0004-2891-5058
** Prof. Dr., Çukurova University, Faculty of Health Sciences, Department of Nursing, Adana, Türkiye.
E-mail: skaradag201778@gmail.com ORCID https://orcid.org/0000-0003-0604-8008

Introduction

Diabetes mellitus has become a global public health concern of the 21st century due to its ever-increasing prevalence[1]. Diabetes mellitus is a chronic metabolic disorder characterized by insufficient production of insulin in the pancreas or the body's inability to utilize insulin effectively[2]. The high preventable morbidity and mortality rates related to diabetes mellitus and its complications, and the associated economic burden cause this disorder to turn into a major health problem[1].

Today, digital health technologies play a crucial role in the treatment, care, and education for patients suffering from diabetes mellitus[3]. Artificial intelligence (AI), which has recently come to the forefront among digital health technologies, refers to application systems or machines capable of imitating human cognitive functions and enhancing their performance with the data they acquire[4]. In terms of its theoretical framework, AI includes various sub-fields such as machine learning algorithms, artificial neural networks, random forest algorithms and deep learning, as well as approaches such as logistic regressions and decision trees[5,6]. AI has a self-evolving structure that learns from past data. Therefore, it is capable of continuously updating and optimizing itself to achieve better outcomes[7]. Researchers have been studying on design and development of AI-based clinical decision support systems ever since the mid-20th century[1].

AI applications provide doctors with a comprehensive clinical decision support system by analyzing a wide range of sources such as medical images, laboratory results, medical histories and current data of patients. Algorithms such as convolutional neural networks can detect diseases from complex datasets, enable early detection of medical conditions by continuously monitoring data of patients, and thereby improve care processes.

Each patient's genetic, biochemical, and clinical conditions vary; therefore, individualized treatment plans are essential. AI analyzes separately genetic profiles, lifestyle, symptoms, and other data of each patient, provides personalized healthcare solutions, and makes this process more effective.

AI-based chatbots inform patients by answering their health-related questions and serve as patient advisors through their user-friendly interface. These chatbots support patients' self-management skills by analyzing their data on nutrition, exercise, sleep patterns, and stress levels, and accordingly giving them personalized recommendations. Thus, they are able to promote their health and elevate their well-being levels.

For healthcare professionals, it provides up-to-date and in-depth information on treatment methods and drug efficacy by processing medical images and analyzing medical literature, accelerates doctors' prescription processes, and increases accuracy rate. For physicians, AI organizes medical records of patients, thus allowing to quickly access their medical histories and facilitating data management and prognosis monitoring[3,8,9].

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

This review aims to both demonstrate how AI can be utilized in management of diabetes mellitus including the advantages it offers and to serve as a guideline for future studies on this field.

Artificial Intelligence in Management of Diabetes Mellitus

Modern medical approaches on diabetes mellitus involve education of patients on diabetes mellitus, regular monitoring of their blood glucose levels, healthy diet, exercise, lifestyle changes, and drug treatment (such as oral antidiabetics and insulin)[10]. Traditional treatment approaches in the prevention and management of diabetes mellitus remain inadequate in various aspects. Patients with diabetes mellitus should be checked regularly to monitor their blood glucose levels and prevent possible complications. An integrated team of nurses, podologists, dieticians, nephrologists, endocrinologists, and ophthalmologists take charge in managing diabetes mellitus. However, this process may be interrupted due to short appointment times, high number of patients, the need for continuous follow-up, insufficient capacity of health services, and lack of healthcare professionals[1,11]. It is considered that AI will help address these issues and mitigate the disease burden of diabetes mellitus in the future. AI is also expected to enhance diabetic patients' quality of life and make the healthcare system run more efficiently in management of this disease[1]. These patients should regularly monitor their blood glucose levels, adhere to appropriate diet and exercise programs, and take their medications properly[12]. Continuous information and support during this management process helps boost patients' motivation and control over this disease[13]. The use of AI-based digital health technologies for managing this disease can help develop diabetes prevention strategies for high-risk populations, follow up patients with diabetes mellitus who are unable to attend physician appointments, keep them informed about their disease, promote self-management, and save time and cost through remote patient monitoring[1]. The disruption in the follow-up of these patients during the COVID-19 pandemic has shown how vital such technologies are for their remote follow-up through virtual clinics[14].

AI can provide data-oriented, sensitive treatment and care methods in addition to existing methods, resulting in a paradigm shift in management of diabetes mellitus[15]. AI-based virtual assistants offer people 24/7 service and answer their questions instantly, send reminders, and provide educational materials. They play a crucial role particularly in monitoring their blood glucose levels and reminding them to take their medications[16].

Artificial Intelligence Applications Used in Management of Diabetes Mellitus

Treatment of diabetes mellitus mostly depends on self-management of the patient. As AI has advanced, patients can generate data for their parameters and manage their diabetes mellitus. Web-based programs can be integrated into smartphone applications and wearable technological devices, making them easier to use[9,17]. AI is capable of analyzing the data of patients and offering personalized care for them; thus, providing an individual-centered approach[18].

A study conducted with patients who suffered from type 1 diabetes mellitus reported that while using an insulin pump, the AI-based virtual nurse support lowered the hemoglobin A1c (HbA1c) levels of patients and reduced number of their hospitalizations[19]. Likewise, another study reported that the virtual assistant application developed for older patients with Type 2 diabetes mellitus improved their self-care, medication intake, physical activity and dietary adherence in such a way that healthcare professionals want them to do[20].

In an AI model developed to predict the development of type 2 diabetes mellitus, the variables of fasting plasma glucose, HbA1c, triglycerides, BMI, gamma-glutamyl transferase, uric acid, age, gender, smoking, alcohol, physical activity, and family history were used. The studies proved that performance of this prediction model was highly effective[21].

Nowadays, AI systems that deliver insulin through an insulin pump in accordance with the values in the continuous glucose measurement system are used for type 1 diabetes mellitus[15]. Likewise, an AI-based system adapted for remote adjustment of insulin dosage sends the data of patient's self-blood glucose measurement to a cloud server, and then a physician can inform patients and offer them advices[22].

Related studies have indicated that algorithms such as deep learning and machine learning can predict blood glucose levels[23]. One AI-based continuous glucose monitoring smartphone application demonstrated that AI could predict hypoglycemic episodes half an hour before they happened and notify the patient in advance using the data with the accuracy of 98.5%.

Advanced AI algorithms, such as decision trees, can detect diabetic foot ulcers in images and assess the risk of amputation[25,26]. The application was developed by Yap et al., to determine whether or not an image contains a diabetic foot ulcer and it has been highly beneficial for clinicians or visually impaired patients in regions where people are not highly trained on management of diabetic foot ulcer[25].

Machine learning and deep learning have shown advances in the field of medical imaging. Image and pattern recognition has become increasingly important in the analysis of retinal scans for individuals with diabetes mellitus. Some applications perform well[27].

Management of diabetes mellitus requires physical health as well as psychosocial support. Depression and anxiety are prevalent among diabetic patients and only complicate the management of the disease. AI-assisted chatbots provide emotional support to patients, boost their motivation, and help them cope with stress. These systems also support regular follow-up and treatment processes by facilitating users' access to healthcare professionals[28].

AI algorithms can provide long-term health monitoring by collecting personal data of patients. Changes in their health conditions can be detected at an early stage, whereby healthcare professionals, especially nurses, adjust their treatment plans in a timely and effective manner[29]. Continuous and individualized support helps patients comply with treatment and keep their blood glucose levels under control; thus, enhancing their quality of life and ability to cope with diabetes mellitus[30].

Artificial Intelligence and Nursing

The impact of AI has been increasing in various roles of nursing, including care, education, counseling, research, management, and collaboration. Tasks such as organizing shift schedules and workflow charts, examining diagnostic and observation procedures, digitally recording patient data, measuring vital signs, and analyzing data of vital signs—all performed with the help of AI—reduce the workload on nurses, allowing them to devote more time to healthcare and carry out these tasks more quickly and accurately. In high-risk healthcare environments (such as radiation, pressure, and chemotherapy rooms), the use of AI-based systems and robots can reduce the frequency of nurses' exposure to hazardous conditions, contribute to improved care standards, and lead to higher job satisfaction and lower turnover rates. In regions offering limited access to healthcare, AI can enable nurses to maintain uninterrupted communication with patients regardless of location, thereby contributing to the equitable and fair delivery of nursing services[31,32].

Concepts such as AI-driven data analysis, digital data security, conscious use, and digital literacy bring new responsibilities for nurses. Nursing educators, nursing managers, and nurses must improve their skills and serve as guides to increase the use of AI technologies in nursing[31].

Failing to account for the balance between humans and technology in decision-making processes, and neglecting the unique interactions nurses have with patients as well as lack of AI support in their roles of providing empathy and emotional support can undermine the human-centered care approach and negatively affect the patient experience[32].

Moreover, the use of AI is not equal across the board because many patients lack digital literacy or cannot use, access, or afford computer technology or the internet. Various strategies, such as improved access and more training may be needed to solve this issue[11].

Furthermore, the widespread use of AI brings along some ethical and security concerns. Security of user data, ethics, and user privacy are critical for the development of these technologies[33]. Also, inaccurate assessments or recommendations made by AI lead to inadequate care; however, who is responsible for that is unclear. Healthcare professionals should remember that AI is only a supportive tool and does not substitute their own clinical judgement[26]. Mechanisms that would supervise the recommendations of these systems should be established and final decisions should be designed to preserve the autonomy of healthcare professionals and maintain nurse-patient interaction and communication[32].

Conclusion

AI has the potential to improve the care of individuals with diabetes mellitus in the healthcare process by offering advantages such as enhancing screening and diagnostic processes, delivering earlier and targeted treatments, predicting possible complications, and lowering morbidity and mortality rates. It is also capable of optimizing care of diabetes mellitus by delivering personalized, precise, and data-oriented support to patients and healthcare professionals. By overcoming challenges and using opportunities effectively, AI could revolutionize diabetes care and improve the lives of millions of people with diabetes mellitus around the world. To maximize the potential of AI, developers, clinicians, and policymakers should collaborate with each other and establish standards for patient-centered design, ethical and regulatory frameworks, along with data security and privacy. Furthermore, healthcare professionals should be trained for the effective integration of AI into clinical practice, and the effectiveness of these interventions should be evaluated with evidence-based, long-term studies.

Author Contributions

Idea / Concept: GY, Design: GY, SK, Supervision/Consultancy: SK, Analysis / Interpretation: GA, SK, Literature Review: GY, Writing the Article: GY, SK, Critical Review: SK.

Funding: No funding has been received.

Conflict of interest: The authors have declared that there is no conflict of interest among them.

REFERENCES

1. Guan Z, Li H, Liu R, et al. Artificial intelligence in diabetes management: Advancements, opportunities, and challenges. *Cell Rep Med*. 2023;4(10):101213.

2. Cleveland Clinic. Diabetes: what It Is, Causes, Symptoms, Treatment & Types. 2023; Available from: https://my.clevelandclinic.org/health/diseases/7104-diabetes

3. Ceviz A, Özden G. Diyabet yönetiminde yapay zeka destekli sanal hemşirelerin rolü ve geleceği. *Artuklu Health*. 2025;11:80-87.

4. McCarthy J, Minsky ML, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Mag*. 2006;27:12–14.

5. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230-43.

6. Li J, Huang J, Zheng L, Li X. Application of artificial intelligence in diabetes education and management: present status and promising prospect. *Frontiers in Public Health*. 2020;8:2296-2565.

7. İşcan H, Durgun Kaygısız A. Yapay zekâ: alt dalları ve uygulama alanları. *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 2024;16(4):201-234.

8. Hoşgör H, Güngördü H. A Qualitative study on the areas of use of artificial intelligence in healthcare. *European Journal of Science and Technology*. 2022;35:395 407.

9. Sarma AD, Devi M. Artificial intelligence in diabetes management: transformative potential, challenges, and opportunities in healthcare. *Hormones*. 2025;24:307–322.

10. Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for The Study of Diabetes (EASD). *Diabetes Care*. 2018;41(12):2669–2701.

11. Mackenzie SC, Sainsbury CAR, Wake DJ. Diabetes and artificial intelligence beyond the closed loop: a review of the landscape, promise and challenges. *Diabetologia*. 2024;67:223–235.

12. Powers MA, Bardsley J, Cypress M, et al. Diabetes self-management education and support in type 2 diabetes: A joint position statement of the American Diabetes Association, the American Association of diabetes educators, and the Academy of nutrition and dietetics. *Clinical Diabetes*. 2017;43(1):40–53.

13. Polonsky WH, Fisher L, Hessler D, Edelman SV. Investigating hypoglycemic confidence in type 1 and type 2 diabetes. *Diabetes Technology and Therapeutics*. 2017;19(2):131–6.

14. Garg SK, Almurashi AM, Rodriguez E. Virtual clinics for diabetes care. *Diabetes Technology & Therapeutics*. 2023;25(1):2–14.

15. Papazafiropoulou AK. Diabetes management in the era of artificial intelligence. *Arch Med Sci Atheroscler Dis*. 2024;25;9:e122-e128.

16. Gal RL, Raghinaru D, Aleppo G, et al. 149-LB: Telemedicine intervention impact on diabetes self-management—Twelve-Month Virtual Diabetes Specialty Clinic (VDISC) Study Outcomes. *Diabetes*. 2023;72(1).

17. Rollo ME, Aguiar EJ, Williams RL, et al. eHealth technologies to support nutrition and physical activity behaviors in diabetes self-management. *Metab Syndr Obes*. 2016;9:381-90.

18. Mahmoudi H, Moradi MH. The progress and future of artificial intelligence in nursing care: a review. *The Open Public Health Journal*. 2024;17:e18749445304699.

19. Mumford B, Oldham V, Lee D, Jones J, Das G. The effectiveness of running virtual clinics as part of insulin pump services for patients with type 1 diabetes. *Endocrine and Metabolic Science*. 2021;3:100083.

20. Buinhas S, Cláudio AP, Carmo MB, et al. Virtual assistant to improve self-care of older people with type 2 diabetes: First prototype. *Communications in Computer and Information Science*. 2019;1016:236–248.

21. Deberneh HM, Kim I. Prediction of type 2 diabetes based on machine learning algorithm. *Int J Environ Res Public Health*. 2021;18:3317.

22. Nimri R, Oron T, Muller I, et al. Adjustment of insulin pump settings in type 1 diabetes management: advisor pro device compared to physicians' recommendations. *J Diabetes Sci Technol*. 2022;16:364-72.

23. Kozinetz RM, Berikov VB, Semenova JF, Klimontov VV. Machine learning and deep learning models for nocturnal high- and low-glucose prediction in adults with Type 1 Diabetes. *Diagnostics*. 2024;14:740.

24. Abraham SB, Arunachalam S, Zhong A, Agrawal P, Cohen O, McMahon CM. Improved real-world glycemic control with continuous glucose monitoring system predictive alerts. *J Diabetes Sci Technol*. 2021;15:91-7.

25. Goyal M, Reeves ND, Rajbhandari S, Yap MH. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. *IEEE J Biomed Health Inform*. 2019;23(4):1730-1741.

26. Persson V, Lovén Wickman U. Artificial intelligence as a tool for self-care in patients with Type 1 and Type 2 Diabetes—An integrative literature review. *Healthcare*. 2025;13(8):950.

27. Skandarani Y, Jodoin PM, Lalande A. GANs for medical image synthesis: an empirical study. *Journal of Imaging*. 2023;9(3):69.

28. McCoy MA, Theeke LA. A systematic review of the relationships among psychosocial factors and coping in adults with type 2 diabetes mellitus. *International Journal of Nursing Sciences*. 2019;6(4):468–477.

29. Wu Y, Yao X, Vespasiani G, et al. Mobile app-based interventions to support diabetes self-management: A systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. *JMIR MHealth and UHealth*. 2017;5(3):e6522.

30. Rony MKK, Parvin MR, Ferdousi S. Advancing nursing practice with artificial intelligence: Enhancing preparedness for the future. *Nurs Open*. 2024;11(1):10.

31. Gökçen Gökalp M, Üzer MA. Nursing Care in the Age of Artificial Intelligence. *SBU Journal of Nursing*. 2024; 6(1), 89-94.

32. Nurkalem S, Gülseven Karabacak B. Artificial intelligence in nursing: The ethical and humanistic dimensions of the technological revolution. *Journal of Health and Life Sciences*. 2025;7(2), 116-123.

33. Almalawi A, Khan AI, Alsolami F, Abushark YB, Alfakeeh AS. Managing security of healthcare data for a modern healthcare system. *Sensors*. 2023;23(7):3612.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER), 29, Special Issue on Artificial Intelligence in Health Applications, (2026): 187-196.

The Healthcare Team in the Age of Artificial Intelligence: Digital Transformation in Healthcare Service

Ramazan DEMİRER*

Abstract

Artificial intelligence (AI) is supporting clinical decision-making processes in healthcare systems, personalizing patient care, and optimizing workflows. The aim of this narrative review is to examine the role of AI technologies in supporting physicians, nurses and physiotherapists, as well as to evaluate the barriers, ethical issues, and educational transformations associated with integration. Articles related to the topic were selected from Google Scholar, PubMed, and Scopus search databases using the keywords "Artificial intelligence, healthcare team, healthcare services, informatics," without any restrictions on publication year. Findings show that clinical decision support systems increase diagnostic accuracy and, through personalized treatment planning, enhance treatment efficacy. In nursing, AI-supported monitoring systems improve patient safety while reducing administrative burden; in physiotherapy, robotic devices, wearable sensors, and machine learning-based movement analysis support rehabilitation. AI-based health education requires new competencies such as health, data, and human literacy. Key barriers include lack of algorithmic transparency, data privacy concerns, bias, and legal uncertainty regarding accountability. In conclusion, AI functions as "augmented intelligence" that complements rather than replaces healthcare professionals. Effective integration requires transparent infrastructures, clear legal boundaries, workforce training, and human-centered practices. Healthcare teams utilizing these AI-supported systems can maximize patient health outcomes.

Keywords: Artificial intelligence, healthcare team, healthcare services, informatics.

Yapay Zekâ Çağında Sağlık Ekibi: Sağlık Hizmetlerinde Dijital Dönüşüm

Öz

Yapay zeka (YZ), sağlık sistemlerinde klinik karar verme süreçlerini desteklemekte, hasta bakımını kişiselleştirmekte ve iş akışlarını optimize etmektedir. Bu geleneksel derlemenin amacı, YZ teknolojilerinin hekimler, hemşireler ve fizyoterapistleri desteklemedeki rolünü incelemek, ayrıca entegrasyon sürecindeki engelleri, etik sorunları ve eğitimsel dönüşümleri değerlendirmektir. Yayımlanma yılı sınırlaması olmaksızın Google Akademik, PubMed ve Scopus arama tabanlarından "Artificial intelligence, healthcare team, healthcare services, informatics" anahtar kelimeleri kullanılarak konu ile ilişkili makaleler ele alınmıştır. Bulgular, klinik karar destek sistemlerinin tanı doğruluğunu artırdığını ve kişiye özel tedavi planlaması sayesinde tedavi etkinliğini yükselttiğini göstermektedir. Hemşirelikte YZ destekli izleme sistemleri hasta güvenliğini iyileştirirken idari yükü azaltmakta; fizyoterapide ise robotik cihazlar, giyilebilir sensörler ve makine öğrenimi tabanlı hareket analizi rehabilitasyonu desteklemektedir. YZ tabanlı sağlık eğitimleri, sağlık, veri ve insan okuryazarlığı gibi yeni yeterlilikler gerektirmektedir. Temel engeller arasında algoritma şeffaflığının olmaması, veri gizliliği endişeleri, önyargı ve sorumlulukla ilgili hukuki belirsizlik yer almaktadır. Sonuç olarak YZ, sağlık profesyonellerinin yerini almayan, onları tamamlayan bir "artırılmış

Derleme Makale (Review Article)
Geliş / Received: 31.03.2026 & **Kabul / Accepted:** 06.07.2026
DOI: https://doi.org/10.38079/igusabder.1920424
* Res. Asst., Istanbul Gelisim University, Faculty of Health Sciences, Nursing Program, Istanbul, Türkiye.
E-mail: rdemirer@gelisim.edu.tr ORCID https://orcid.org/0000-0002-5660-8360

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

zeka” işlevi görmektedir. Etkin bir entegrasyon için şeffaf altyapılar, net yasal sınırlar, iş gücü eğitimi ve insan merkezli uygulamalar gereklidir. Bu sistemleri kullanan sağlık ekipleri, hasta sağlık sonuçlarını maksimize edebilecektir.

Anahtar Sözcükler: Yapay zeka, sağlık ekibi, sağlık hizmetleri, bilişim.

Introduction

Artificial intelligence (AI), in general terms, is the ability of computer systems or machines designed to simulate human intelligence by performing tasks that would normally require human cognition, such as learning, reasoning, and problem-solving[1]. Today, AI is deeply rooted in finance, commerce, and science; for example, Amazon's product recommendation systems, transportation networks like Uber, and intelligent personal assistants like Siri or Alexa have become an integral part of life[2-5]. In healthcare, AI is having a revolutionary impact by leveraging the power of big data to support evidence-based clinical decision-making, personalize patient care, and digitally transform healthcare processes[4,6,7]. Current applications are beginning to permeate every aspect of healthcare, from medical image analysis and drug research to home management of chronic diseases and surgical robotics systems[4].

Advanced AI technologies and tools are being integrated into the healthcare ecosystem. These technologies include machine learning (ML), which makes predictions by identifying patterns in data, and deep learning (DL), which uses complex neural networks[2,4,8]. Another key component, natural language processing (NLP), has the ability to extract critical information by analyzing unstructured text data such as physician notes and laboratory reports[4,6]. In addition, robots that assist in surgical operations, exoskeletons used in rehabilitation processes, chatbots that interact with patients, and wearable health devices (WHDs) that continuously monitor vital signs are independent technological tools available to healthcare professionals[8,9]. These systems transform traditional "record and storage" systems into active "intelligence systems," providing healthcare teams with an advanced work infrastructure[4].

The outcomes of AI interventions in healthcare services are operational efficiency and team member performance indicators. The success of AI applications is measured by increased diagnostic accuracy, early detection of diseases (sepsis, cancer, etc.), and a reduction in medical error rates[2,6]. In addition, a significant process outcome is the reduction of administrative burden on healthcare personnel and the increase in time dedicated to direct patient care through the automation of routine tasks[2,4,6]. Furthermore, it includes lowering healthcare costs, improving patient satisfaction and clinical safety, and optimizing resource allocation (such as personnel and equipment planning). In fields such as physiotherapy and nursing, patient adherence to treatment and improvements in self-management skills are critical outcomes directly impacted by the use of AI[6,10].

"Augmented intelligence" that complements human intelligence, rather than replacing it. AI's positioning as a partner in the field of intelligence lies in its ability to support the limited cognitive capacity of healthcare professionals by analyzing massive datasets[4,8,9]. AI enhances clinical effectiveness, safety, and accessibility, enabling value-based care[4,10]. In the future, healthcare professionals will have the opportunity to focus more on unique human skills such as empathy, persuasion, and complex case management, moving beyond routine and administrative tasks thanks to AI[6,9]. Given the speed of technological advancement, it is vital for the sustainability of modern healthcare systems that healthcare leaders and teams adapt to this transformation, gain data literacy, and integrate AI into clinical practice within an ethical framework[2,4,9,10].

This study is a narrative review that comprehensively examines the use of artificial intelligence technologies in healthcare services and the implications of this transformation for the healthcare team[1]. Google Scholar, PubMed, and Scopus search databases were used during the literature review process; articles were selected using the keywords "Artificial intelligence, healthcare team, healthcare services, informatics," without any restrictions on publication year. Within the scope of the review, by focusing on the roles of physicians, nurses, and physiotherapists as the primary stakeholders of the healthcare team; the contributions of AI to these professional groups, the ethical and legal barriers encountered, and the requirements for educational transformation have been analyzed and synthesized[1,2].

Fundamentals of Artificial Intelligence Technologies and Their General Contribution to Health Care Services

Artificial intelligence refers to computer systems that simulate elements of human intelligence such as learning, reasoning, and problem-solving[4,11]. The evolution of AI in healthcare has accelerated thanks to increasing computing power, the availability of massive datasets, and improved algorithms[4,8]. Today, AI is making its impact felt in every aspect of healthcare[4,12-15]. The main AI technologies used in healthcare are:

Machine Learning (ML)

Machine Learning is defined as "the ability of a machine to mimic intelligent human behavior." An algorithm is trained to learn from data and then make decisions based on similar features or variables from new data. The intersection of various disciplines, particularly mathematics, statistics, and computer science, is a crucial component of data science required to implement various ML models. Despite the power of human intelligence, humans are prone to error due to their limited short-term memory[1]. The increase in data volume, coupled with the improved capacity of computers to analyze this data and develop various machine learning models, creates the possibility of supporting individuals' decision-making processes with rich contextual information by leveraging machine learning. Today, machine learning is beginning to be applied in many different medical fields, such as monitoring heart failure, clinical decision support systems, and medical imaging[1,4,16].

Natural Language Processing (NLP)

NLP is a subfield of artificial intelligence that deals with the use of computers to process and interpret data in natural language. Applications of NLP include translation, information extraction, summarization, topic modeling, sentiment analysis, translating handwritten text to text and vice versa, and answering questions[17]. These skills improve decision-making by transforming vast medical literature and patient histories into concise summaries, reducing the burden of documentation and manual data entry for doctors and nurses[18]. Despite challenges such as model bias and implementation costs, NLP integration can provide personalized care, precision medicine, and operational efficiency to healthcare systems[18,19]. A study by Thakur et al. (2024) found that NLP successfully extracted valuable medical information from disorganized health data, improving diagnostic accuracy and accurately predicting treatment outcomes[20].

Deep Learning (DL)

DL involves using a neural network with many layers (deep structure) between input and output, and its greatest advantage is that it is data-driven, highly representative, and can automatically learn hierarchical features and perform feature extraction and classification on a single network. DL can be used to model or simulate an intelligent system or process using labeled training data. Recently, deep learning (DL) has been widely used in medical applications such as anatomical modeling, tumor detection, disease classification, computer-aided diagnosis, and surgical planning[21]. A study by Jin et al. (2023) concluded that deep learning has the potential not only to diagnose COVID-19 but also to assess disease progression and prognosis, suggest treatment plans, and assist federal governments in formulating intelligent steps to control and prevent the spread of the disease[22].

Specific Contributions to Healthcare Team Members

Medical and Clinical Decision Support Processes

Provide significant support to healthcare professionals in the early diagnosis of diseases, treatment planning, and patient management[23]. AI accelerates and refines physicians' diagnostic and treatment decisions. Studies have shown that AI algorithms can perform at the same level as humans, and in some cases even better, in the analysis of radiological images (CT, MRI, X-ray)[8,24]. Studies have proven that AI can perform better than radiologists in breast cancer screening[3,4]. AI-based algorithms offer strong support to healthcare professionals in the early diagnosis, treatment planning, and patient management of cardiovascular diseases; however, these systems are not the ultimate decision-makers, but rather auxiliary tools[23].

Nursing Services and Care Process

Nursing Services and Care Process has become a significant resource in clinical decision-making, patient monitoring, and workflow optimization[6]. The integration of AI solutions

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

into nursing practice provides significant clinical and service gains. Thanks to AI-powered monitoring tools such as wearable technologies and real-time notification systems, nurses can detect subtle physiological changes, such as fever or pain signals, much earlier than traditional methods. This makes it possible to reduce complications, shorten hospital stays, and lower readmission rates. Indeed, numerous studies have shown that early warning algorithms enable clinical teams to react more quickly, improving patient safety and treatment outcomes. From a service perspective, AI-based automation of routine tasks such as scheduling, administrative record keeping, and predictive workload classification has enabled more efficient resource allocation. This increase in efficiency has resulted in a measurable decrease in burnout and a significant increase in job satisfaction among nurses. This is because nurses are now able to dedicate more time to direct patient care[25]. In recent years, innovative technologies such as artificial intelligence (AI) have revolutionized wound care by offering more effective and efficient solutions for acute and chronic injuries[26]. Mobile applications and image recognition systems can track the size and healing process of wounds, such as diabetic foot ulcers, with similar precision to a professional wound care nurse. Studies have shown that AI algorithms increase diagnostic accuracy, accelerate healing, reduce pain, and improve cost-effectiveness in wound care[27]. On the other hand, while AI can support documentation and diagnostic processes, its current limitations in relational, sensory, and adaptive aspects have raised concerns among nurses regarding its suitability for wound care in a home environment. Successful AI integration must account for the realities of nursing practice and enable technological tools to enhance the bodily, relational, and adaptive dimensions of wound care[28]. AI-based systems enable remote monitoring of patients with conditions such as atrial fibrillation and improve treatment adherence[6]. Clinical staff spend almost half their day on electronic health records and administrative tasks. This reduces the time spent directly with patients and increases burnout[29]. NLP systems allow nurses to dedicate more time to direct patient care by reducing the time spent on administrative tasks[10].

Physiotherapy and Rehabilitation Practices

Today, artificial intelligence technologies are used in multidisciplinary health research fields. Physiotherapy and Rehabilitation is one of these fields. AI is used to improve the patient care process by assisting physiotherapists in conducting comprehensive assessments, predicting patient performance, or making diagnoses[30]. Furthermore, research has revealed many more uses of artificial intelligence in medical and rehabilitation applications, such as problem-solving, X-ray diagnosis, planning treatment protocols, and physical manipulation of patients[10]. All these functions of artificial intelligence are fundamental elements of physiotherapy professional practice. Studies have shown that AI-supported applications can perform video analysis of gait in individuals with Parkinson's and osteoarthritis, automating the diagnosis of abnormalities in these patients[9,24]. Various studies have shown that AI-supported application tools can predict injury risks by calculating lower extremity joint angles[10,24]. AI-supported rehabilitation robots and exoskeletons are widely used in post-stroke rehabilitation or to assist individuals with physical disabilities in their daily activities, facilitating rehabilitation processes[8]. Furthermore, through wearable technologies, smartwatches, and inertial sensors, exercise compliance and performance of patients undergoing rehabilitation can be monitored with a high accuracy of 99.4%[9].

Artificial Intelligence Transformation in Health Education

The possibilities offered by artificial intelligence promise to overcome traditional educational models known for their rigid structures, enabling seamless lifelong learning and easy adaptation to rapidly changing healthcare environments[31]. Integrating AI into healthcare education not only improves the effectiveness and quality of education but also innovates the ways in which health-related knowledge is shared with others; ultimately, this equips current and future healthcare teams with the skills needed to address a wide range of health challenges[32]. The integration of AI into healthcare services also necessitates fundamental changes in the training of healthcare professionals. Healthcare professionals are now expected to possess three core literacies. These are data literacy, technical literacy, and human literacy[10]. Studies have shown that AI-powered chatbots in nursing and medical education can improve students' knowledge levels and clinical decision-making abilities more effectively than traditional methods[6]. Virtual reality (VR) and augmented reality (AR) allow medical students to learn anatomy without the need for cadavers or to experience complex surgical procedures without risk[8]. One of the transformative effects that AI brings to medical education is its ability to provide feedback and evaluation in real time. Traditional examination and evaluation methods are both time-consuming and do not always accurately reflect the student's true competencies. In contrast, AI-powered assessment tools provide more objective and faster feedback by instantly analyzing the student's performance[33]. This increases the student's learning effectiveness.

Barriers to Implementation, Ethical Issues, and Perceptions

Barriers to ımplementation, ethical ıssues, and perceptions of AI, its widespread adoption faces significant obstacles. Firstly, the ambiguity surrounding how algorithms arrive at decisions makes it difficult for physicians to trust the system[2]; secondly, the use of large datasets raises concerns about patients' privacy rights and the risk of cyberattacks[3]. Furthermore, the lack of diversity in training data can lead to AI producing erroneous or discriminatory results for specific demographic groups[2,4,7] while the question of who is responsible in the event of an error in an AI-assisted decision remains legally ambiguous[2]. The general perception of healthcare professionals is mixed; while many find AI useful, a large majority state that they do not fully understand its working principles[3], and the majority believe that AI will be a helpful tool rather than taking away their jobs[3,4]. However, while AI integration reduces workload, it also brings new risks such as the atrophy of clinical skills, leaving complex cases solely to doctors, employee alienation, and increased inequalities[29].

Conclusion

Artificial intelligence (AI) is not a competitor replacing humans in healthcare, but rather an "augmented intelligence" partner that enhances the capabilities of professionals. Studies show that AI has the potential to increase efficiency, quality, and safety in a wide range of areas in clinics, from decision-making support to imaging, rehabilitation, and administrative management. However, AI development alone is not sufficient for this transformation; healthcare leaders and policymakers need to develop strategic plans and create a culture of innovation. One of the biggest obstacles to successfully implementing AI applications in the healthcare system is the lack of transparency in decision-making processes. Therefore, frameworks such as legal responsibility, accountability, and data privacy need to be clarified to ensure employee trust in the system. Furthermore, important factors such as standardization, data quality, and bias prevention need to be carefully managed. National-level strategies should prioritize and develop AI. These strategies are crucial for establishing a system that ensures all stakeholders benefit from this technology equally and safely.

Preparing healthcare professionals for this new era necessitates changes in education systems and curricula. Future healthcare professionals should be equipped not only with clinical knowledge but also with technical literacy, data literacy, and interpersonal literacy. While AI will handle cognitively demanding tasks such as complex analyses and data mining, clinicians will be able to dedicate more time to human skills that AI cannot replicate, such as compassion, empathy, ethical evaluation, and complex case management. In this way, AI will play a role in reducing workload and increasing job satisfaction for all healthcare professionals, from nurses to physiotherapists. In conclusion, the integration of AI into healthcare is an inevitable reality. However, this process must be managed with a human-centered approach. Studies predict that AI systems will not completely replace healthcare professionals, but that professionals working with AI will surpass those who do not. Instead of resisting technological advancements, healthcare team members should combine the opportunities offered by these tools with their clinical knowledge to elevate patient care outcomes to an excellent level. The sustainability of modern healthcare systems depends on the successful synthesis of the technological power brought about by digital transformation and the ancient human values of medicine.

REFERENCES

1. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. *J Am Col Cardiol*. 2017;69(21):2657-2664.

2. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A systematic review of the barriers to the implementation of artificial intelligence in healthcare *Cureus*. 2023;15(10):e 46454. doi: 10.7759/cureus.46454

3. Castagno S, Khalifa M. Perceptions of artificial intelligence among healthcare staff: a qualitative survey study. *Front Artif Intel*. 2020;3:578983. doi: 10.3389/frai.2020.578983

4. Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. *Healthc Management Forum*. 2019;32(6):1-9. doi: 10.1177/0840470419873123

5. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare application. In: Bohr A, Memarzadeh K, eds. *Artificial Intelligence in Healthcare*. Academic Press; 2020:25-60.

6. Martinez-Ortigosa A, Martinez-Granados A, Gil-Hernández E, Rodríguez-Arrastia M, Ropero-Padilla C, Roman P. Applications of artificial intelligence in nursing care: a systematic Review. *J. Nurs Manag*. 2023;2023:3219127. doi: 10.1155/2023/3219127

7. Richardson JP, Smith C, Curtis S, et al. Patient apprehensions about the use of artificial intelligence in healthcare *NPJ Digit Med*. 2021;4:140. doi: 10.1038/s41746-021-00509-1

8. Bohr A, Memarzadeh K. Artificial intelligence in healthcare: a comprehensive review. In: Bohr A, Memarzadeh K, eds. *Artificial Intelligence in Healthcare*. Academic Press; 2020:1-24. doi: 10.1016/B978-0-12-818438-7.00001-0

9. Aggarwal R, Ganvir SS. Artificial intelligence in physiotherapy. *Physiotherapy - J Indian Assoc Physiother*. 2022;15(2):55-57. doi: 10.4103/pjiap.pjiap _3_22

10. Ravali RS, Vijayakumar TM, Lakshmi KS, et al. A systematic review of artificial intelligence for pediatric physiotherapy practice: past, present, and future. *Neurosci Inform*. 2022;2(4):100045. doi: 10.1016/j.neuri .2022.100045

11. Machine learning explained. MIT Sloan. Accessed March 30, 2026. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

12. Saravi B, Hassel F, Ülkümen S, et al. Artificial intelligence-driven prediction modeling and decision performing spine surgery using hybrid machine learning models. *J Pers Med*. 2022;12(4):509. doi: 10.3390/jpm12040509.

13. Connor CW. Artificial intelligence and machine learning in anesthesiology. *Anesthesiology*. 2019;131(6):1346-1359. doi: 10.1097/ALN.0000000000002694

14. Awan SE, Sohel F, Sanfilippo FM, Bennamoun M, Dwivedi G. Machine learning in heart failure: ready for prime time. *Curr Opin Cardiol*. 2018;33(2):190-195.

15. Peiffer-Smadja N, Rawson TM, Ahmad R, et al. machine learning for clinical decision support in infectious diseases a narrative review of current applications. *Clin Microbiol Infect*. 2020;26(5):584-595. doi: 10.1016/j.cmi.2019.09.009

16. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medicine imaging: intelligent imaging. *J Med Imaging Radiat Sci*. 2019;50(4):477-487.

17. McNicholas BA, Madden MG, Laffey JG. natural language processing in critical care: opportunities, challenges, and future directions. *Intensive Resort Med*. 2025;51(3):585-589.

18. Yadav S. Embracing artificial intelligence: revolutionizing nursing documentation for a better future. *Cureus*. 2024;16(4):e 58123. doi: 10.7759/cureus.57725

19. Kumar A, Gond A. Natural language processing: Achieving Healthcare benefits via NLP. *IJNRD*. 2023;8(12):244-252. doi: 10.14293/S2199-1006.1.SOR-.PPMLQWK.v 1

20. Thakur GK, Thakur A, Khan N, Anush H. The role of natural language processing in medical data analysis and healthcare automation. In: *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* . IEEE; 2024:1-5.

21. Chen YW, Jain LC. *Deep Learning in Healthcare: Paradigms and Applications*. Springer; 2020. doi: 10.1007/978-3-030-32606-7

22. Jin S, Liu G, Bai Q. Deep learning in COVID-19 diagnosis, prognosis and treatment selection. *Mathematics*. 2023;11(6):1279. doi: 10.3390/math11061279

23. Bozyel S, Şimşek E, Koçyiğit D, et al. Artificial intelligence-based clinical decision support systems in cardiovascular diseases. *Anatol J Cardiol*. 2024;28(2):74-84.

24. Tack C. Artificial intelligence and machine learning applications in musculoskeletal Physiotherapy. *Musculoskeletal Sci Pract*. 2019;39:164-169.

25. Hassanein S, El Arab RA, Abdrbo A, et al. Artificial intelligence in nursing: an integrative review of clinical and operational impacts. *Front Digit Health*. 2025;7:1552372.

26. Encarnação R, Manuel T, Palheira H, Neves-Amado J, Alves P. Artificial intelligence in wound resort education: protocol for a scoping Review. *Nurs Rep*. 2024;14(1):627-640.

27. Ganesan O, Morris MX, Guo L, Orgill D. A review of artificial intelligence in wound care. *Artif Intell Surg*. 2024;4(4):364-375. doi: 10.20517/ais.2024.68

28. Karnehed S, Larsson I, Petersson L, Erlandsson LK, Tyskbo D. Navigating artificial intelligence in home healthcare: challenges and opportunities in nursing wound care. *BMC Nurs*. 2025;24(1):660. doi: 10.1186/s12912-025-03348-7

29. Pavuluri S, Sangal R, Sather J, Taylor RA. balancing act: the complex role of artificial intelligence in addressing burnout and healthcare workforce dynamics. *BMJ Health Resort Inform*. 2024;31(1):e 101120. doi: 10.1136/bmjhci-2024-101120

30. Vélez-Guerrero MA, Callejas-Cuervo M, Mazzoleni S. Artificial intelligence-based wearable robotics exoskeletons for upper limb rehabilitation: a review. *Sensors*. 2021;21(6):2146. doi: 10.3390/s21062146

31. Wang X, He X, Wei J, Liu J, Li Y, Liu X. Application of artificial intelligence to the public health Education. *Front Public Health*. 2023;10:1087174.

32. Wang J, Li J. Artificial intelligence empowering public health education: prospects and challenges. *Front Public Health*. 2024;12:1389026. doi: 10.3389/fpubh.2024.1389026

33. Sriram A, Ramachandran K, Krishnamoorthy S. Artificial intelligence in medical education: transforming learning and practice. *Cureus*. 2025;17(3):e78901.

IGUSABDER WRITING RULES (Rev.10)

1) **Istanbul Gelisim University Journal of Health Sciences** *(abbreviated title: IGUSABDER)* published three times a year, includes articles on specific research, case report and review related to Health Sciences **(Nutrition and Dietetics, Child Development, Physical Therapy and Rehabilitation, Ergotherapy, Gerontology, Nursing, Audiology, Speech and Language Therapy, Orthotics- Prosthetics, Perfusion, Healthcare Management, Social Service, Clinical, Paraclinical, Biological and Basic Sciences.)** Articles related to the field of dentistry are not included in our journal. The articles submitted to be published should be written in English. A full version of your manuscript submitted to IGUSABDER must not have been previously published.

2) The articles submitted for publication should be prepared in the format of

- Times New Roman style, font size 12, A4 paper size, 1.5 line spacing and 2.5 cm margins of all edges.
- Visual items like figures and tables should be written in the language the article is written and they should be placed appropriately in the text with the necessary explanations.
- The titles of the tables, figures and graphics should be on the top and left aligned.
- The abbreviations used in the article should be stated clearly where it is used for the first time and their abbreviations should be indicated between parantheses and specific abbreviations should not be used.
- The first letters of the main titles should be capital, bold and center aligned.
- The first letters of the sub titles should be capital, bold and left aligned.
- Paragraph indentation should not be used.

3) The submission process of the article and its appendices should be carried out through DergiPark online article submission system at the address of http://igusabder.gelisim.edu.tr or https://dergipark.org.tr/en/pub/igusabder. During the submission, the authors should upload the figures of the manuscript to the online manuscript submission system. First of all, if the manuscript is accepted for publication, the copyright transfer agreement form signed by all the authors should be sent to the editorial office. **"Ethical Statement and Copyright Transfer Form"** should be signed by all authors and should be loaded to the system. Studies that requiring ethics committee approval, the **"Ethics Committee Approval Document"** should also be uploaded to the system. If the article is not accepted, the related forms will be invalid. In case the article is not accepted, related form will be invalid.

4) According to TUBITAK ULAKBIM DergiPark rules are required, each author should have an ORCID information and should add it to the journal profile. Authors need to enter their ORCID identity into the system while uploading their article. It is possible to obtain the ORCID number free of charge through the https://orcid.org/signin website.

5) Authors should indicate the institution, date and approval number they have received in the **Ethics Committee Approval** required for the article they wish to publish, at the end of the Discussion section of the Materials and Methods section. Statement of **"Informed Volunteer Consent/Approval Form"** to participants should be included at the end of the discussion section in statements such as conflicts of interest, thanks, supporting organizations and case reports. This information should be written in text language.

6) For all articles, Turkish and English abstracts

- Should be no more than 400 words and they should be written with at least 3 keywords.
- Only the first letter of the first key word should be capital, the other key words should be written with small letters with comas between them.
- Key words in English and Turkish should be selected in accordance with the Turkish Scientific Terms. Accessed from (http://www.bilimterimleri.com/).
- The necessary changes recommended for authors who do not have access to the Turkish Scientific Database are made by the Editorial Office.

7) ARTICLE TYPES

7.1. Original Research Articles: Original (full-length) Articles are original and proper scientific papers based on sufficient scientific research, observations and experiments. Articles should consist of title, abstract and keywords in Turkish and title, abstract and keywords in English as well as Introduction, Material & Methods, Results, Discussion, Conclusion and References parts. Also it should not exceed 12 pages except in exceptional circumstances (including text, tables and illustrations). There is no limit for the number of references.

The abstract should include the aim, method, results and the conclusion and it should be written accordingly with the example given below.

Example:

Abstract

Aim: The research has been made descriptively in order to determine the levels of the communication skills and the related variables.
Method: The universe of the research consists of 1116 students at the School of Health Sciences of a private university. In the research the universe has not been selected and the universe consists of 615 students that has accepted to join the research. The information form and communication skills scale has been used to collect the data. The data has been evaluated with the SPSS programme.
Results: According to the research findings, the communication skills scale score avarage is 156.1± 13.5. When the relationship between the sociodemographic characteristics and the communication skills scale and the sub dimensions score avarage is analyzed, in women behavioral sub dimension score avarage is higher at students that have taken a theoretical education about communication ($p<0.05$). The communication skills scale of the students' whose father's education levels are literate is higher ($p<0.05$)
Conclusion: As a result of the research it has been determined that the communication skills score average is at medium level and it can be suggested that more lessons about communication skills should be given at all departments of the School of Health Sciences.

Keywords: Health, student, communication.

Örnek:

Öz

Amaç: Araştırma, Sağlık Bilimleri Yüksekokulu öğrencilerinin iletişim becerileri düzeylerini ve ilişkili değişkenleri belirlemek amacıyla tanımlayıcı olarak yapılmıştır.

Yöntem: Araştırmanın evrenini, İstanbul'da bulunan bir özel üniversitenin Sağlık Bilimleri Yüksekokulu'nda öğrenim gören 1116 öğrenci oluşturmuştur. Araştırmada örneklem seçimine gidilmeksizin çalışmaya katılmayı kabul eden 615 öğrenci çalışma kapsamına alınmıştır. Verilerin toplanmasında, Bilgi Formu ve İletişim Becerileri Ölçeği kullanılmıştır. Veriler SPSS programı ile değerlendirilmiştir.
Bulgular: Araştırma bulgularına göre; iletişim becerileri ölçeği puan ortalaması 156,1±13,5 bulunmuştur. Öğrencilerin sosyodemografik özellikleri ile iletişim becerileri ölçeği ve alt boyutlarının puan ortalamaları ilişkisi değerlendirildiğinde; kadınlarda, odyoloji bölümünde okuyanlarda ve iletişim ile ilgili teorik eğitim alanlarda davranışsal alt boyutu puan ortalaması daha yüksek bulunmuştur ($p<0,05$). Baba eğitim düzeyi okuryazar olan öğrencilerin iletişim becerileri ölçeği puan ortalaması daha yüksek bulunmuştur ($p<0,05$).
Sonuç: Araştırma sonucunda iletişim becerileri puan ortalaması orta düzeyde olduğu saptanmış olup, iletişim becerilerinin daha da geliştirilebilmesi için Sağlık Bilimleri Yüksekokulunun tüm bölümlerinde iletişim becerileri ile ilgili derslere daha fazla yer verilmesinin faydalı olacağı düşünülmektedir.

Anahtar Sözcükler: Sağlık, öğrenci, iletişim.

7.2. Case Report: These are the articles that describe rare significant findings encountered in the application, clinic and laboratory of related fields. The reports should include the sections of Introduction, Case History, Conclusion and References and they should not exceed 6 pages. It should be declared that the "Informed Volunteer Consent / Approval Form" was signed at the end of the discussion section.

7.3. Review: These are original articles that the author reviews a current and significant subject through the results that the author obtains from his/her own point of view and research. The reviews should include the sections of Introduction, Conclusion and Suggestions and References and they should not exceed 12 pages.

8) Author/Authors' e-mail addresses, institutional information, ORCID information, main text file must be included as footnotes on the first page and added to relevant places in the system during online application. In addition, in research articles -as in the example- ETHICAL STATEMENT note should be included in this part. The language of the information must be the same as the language of the article. Here is an example of how the format of this information is:

(Original Research Article)
Received: .... & /Accepted: ....
* Lecturer, Istanbul Gelisim University, Vocational School of Health Sciences, Pathology Laboratory Techniques Program, Istanbul, Türkiye. E-mail: ....
ORCID https://orcid.org/.....
** Prof. Dr., Kırıkkale University, Faculty of Arts and Sciences, Department of Biology, Kirikkale, Türkiye. E-mail: .... ORCID https://orcid.org/.........
- -

ETHICAL STATEMENT: This study was carried out with the approval of the Ethics Committee of .... University, dated .../.../........ and numbered .... A signed subject consent form in accordance with the Declaration of Helsinki was obtained from each participant.

9) The necessary descriptive information about article (thesis, project, financial supports etc.) should be explained as footnote in article title.

10) If cited in the text, it should be numbered as superscript. Also, References should be listed with numerical order as they appear in the text and the reference number should be indicated inside the parentheses at the cited text place. (For instance............. has been found[1].)

References should be written by using **Journal of American Medical Association** (JAMA Citation Style). This information can be accessed from the links below. (http://guides.med.ucf.edu/ld.php?content_id=5191991) (https://med.fsu.edu/userFiles/file/AmericanMedicalAssociationStyleJAMA.pdf)

REFERENCES

11.1. BOOKS

11.1.1. One Author	**Author last name Author's first initials. *Title of Book*. Edition number. Place of publication: Publisher; year.** Duyan V. *Sosyal Hizmet: Temelleri, Yaklaşımları, Müdahale Yöntemleri*. Ankara: Nar Yayınevi; 2010. Bickley LS. *Bate's Guide to Physical Examination and History Taking*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2013.

11.1.2. More than one author **(List all authors if six or less, otherwise list three followed by "et al"or "ve ark")**	**Author(s) last name Author(s)' first initials separated by commas. *Title of Book*. Place of publication: Publisher; year.** Tayfur M, Barış O, Nazan Baştaş N. *Diyetisyenlik Eğitimi ve Meslek Etiği*. 2. baskı. Ankara: Hatiboğlu Yayınevi; 2014. Shils M, Shike M, Olson J, Ross AC. *Modern Nutrition in Health and Disease*. 9th ed. Baltimore:Lippincott Williams & Wilkins, 1998.
11.1.3. Edited book	**Author(s) last name Author(s)' first initials, ed(s). *Title of Book*. Edition number. Place of publication: Publisher; year.** Norman IJ, ed. *Mental Health Care for Elderly People*. New York: Churchill Livingston; 1996.

11.1.4. Chapter or article from a book	**Author(s) last name Author(s)' first initials of article. Title of article. In: Editor's name, ed(s). *Title of Book*. Edition number. Place of publication: Publisher; Year.** Cohen M. Chronic and Acute. In: Sapphire P, ed. *The Disenfranchised*. Amityville, New York: Baywood Publishing; 2013. Phillips SJ, Whisnant JP. Hypertension and stroke. In: Laragh JH, Brenner BM, eds. *Hypertension: Pathophysiology, Diagnosis and Management*. 2nd ed. New York: Raven Press; 1995.

11.2. JOURNALS

Author(s) last name Author(s)'s first initials. Article title. *Journal Title*. Year;volume(issue):Inclusive page numbers.(List all authors if six or less, otherwise list three followed by "et al"or "ve ark") Sevinç S, Yavaş Çelik M. Akraba evliliklerinin çocuk sağlığına etkisi ve hemşirelik yaklaşımı. *Sağlık ve Toplum*. 2016;2:23-28. Nabavi SM, Habtemariam S, Daglia M, et al. Neuroprotective effects of ginkgolide B against ischemic stroke: a review of current literature. *Curr Top Med Chem*. 2015;15(21):2222-2232

11.3. ELECTRONIC SOURCES

11.3.1. Electronic articles from online journals with DOI available	**Author(s) last name Author(s)'s first initials. Title of article. Name of Journal. Year;volume(issue):pages. doi:11.1111.** Üstün G, Aluş Tokat M. Gestasyonel diyabet emzirme sonuçları için ne kadar önemli? *Perinatoloji Dergisi*. 2011;19(3):123-129. doi: 10.2399/prn.11.0193005. Rosenbaum M, Leibel RL. Models of energy homeostasis in response to maintenance of reduced body weight. *Obesity*. 2016;24(8):1620-1629. doi: 10.1002/oby.21559.
11.3.2. Electronic articles from online journals without DOI available	**Author(s). Title of article. *Name of Journal*. Year;vol(issue):pages. URL. Published date. Updated date. Accessed date.** Thomas JL. Helpful or harmful? Potential effects of exercise on select inflammatory conditions. *Phys Sportsmed*. 2013;41(4):93-100. https://physsportsmed.org/psm.2013.11.2040. Accessed November 22, 2013.
11.3.3. (eBOOK) Book chapter/ article from eBOOK	**Author(s) last name Author(s)'s first initials of chapter/article. Title of article. In: Editor's name, ed(s). Title of Book. Edition number. Place of publication: Publisher; year. URL. Accessed date: Chapter or page number or section number.** Fields HL, Martin JB. Pain: pathophysiology and management. In: Longo DL, Fauci AS, Hauser SL, Kasper DL, Loscalzo J, Jameson JL, eds. *Harrison's Principles of Internal Medicine*. 18th ed. New York: McGraw-Hill; 2012. http://www.accessmedicine.com.ezproxy.med.ucf.edu/resourceTOC.aspx?resourceID=4. Accessed November 22, 2013:71-73.
11.3.4. Web pages	**Author(s) or responsible body. Title of item cited. Name of website. URL. Published date. Updated date. Accessed date.** World Health Organization. Philippines: Assistance and response after Typhoon Haiyan. World Health Organization. http://www.who.int/features/2013/philippinestyphoonhaiyan/en/index.html. Published November 2013. Accessed November 22, 2013.

11.4. OTHER SOURCES

11.4.1. Thesis	**Author last name Author's first initials. Title of Thesis. [type of thesis]. Name of the place where the thesis was made, Name of the country: Name of the department, Name of the Institute; year.** Undeman C. Fully Automatic Segmentation of MRI Brain Images [master's thesis]. Stockholm, Sweden: NADA, Royal Institute of Technology;2001.
11.4.2. Conference paper	**Author(s) last name Author(s)' first initials. Title of conference paper. In: Title of conference; Day month, year; Name of the place where the conference was made, Name of the country.** Bengtsson S, Solheim BG. Enforcement of data protection, privacy and security in medical informatics. In: Proceedings of the 7th World Congress on Medical Informatics; Sep 6-10, 1992; Geneva, Switzerland. Abstract 209.
11.4.3. Newspaper article	**Author(s) last name Author(s)' first initials. Title of newspaper article. *Name of the newspaper*. Day month, year.** Lee G. Hospitalizations tied to ozone pollution: study estimates 50,000 admissions annually. *The Washington Post*. Jun 21, 2006:A3.

12) The Latin expression such as species names of bacterium, virus, parasite and fungus and anatomical terms must be written in italic character keeping their original forms.

13) The editorial board has the right to perform necessary revision and reduction in the article submitted for publication and to express recommendations to the authors. The articles sent to authors for correction should be sent back to the editorial office within the time specified. Articles which are pre-estimated and deemed suitable for evaluation are sent to at least two referees specialized in the related field and the article can be published after the reports and evaluations come from experts on consultative committee are evaluated and approved.

14) All responsibilities from published articles merely belong to the authors.

15) Once the article has been uploaded to the system, it is mandatory that the plagiarism checker (iThenticate, TURNITIN, URKUND) for the article be uploaded to the relevant section of the "Submit Article" module on the web page of our journal. Except for bibliography and abstract, "exclusion" should not be applied in the similarity report.

16) There is no copyright fee for the authors.

17) Published journal is sent to the authors for free.

18) All persons submitting articles to the journal are deemed to have accepted these rules.