

Biyoistatistik: Teori, Metodoloji ve Uygulama

**Editör
Prof. Dr. Salim G ng r**

Biyoistatistik: Teori, Metodoloji ve Uygulama

Editör

Prof. Dr. Salim G ng r

İmtiyaz Sahibi
Platanus Publishing®

Editör
Prof. Dr. Salim Güngör

Kapak & Mizanpaj & Sosyal Medya
Platanus Yayın Grubu

Birinci Basım
Ekim, 2025

Yayımcı Sertifika No
45813

ISBN
978-625-6454-60-6

©copyright
Bu kitabın yayım hakkı Platanus Publishing'e aittir. Kaynak gösterilmeden alıntı yapılamaz, izin alınmadan hiçbir yolla çoğaltılamaz.

Adres: Natoyolu Cad. Fahri Korutürk Mah. 157/B, 06480, Mamak,
Ankara, Türkiye.

Telefon: +90 312 390 1 118
web: www.platanuspublishing.com
e-mail: platanuskita@gmail.com



Platanus Publishing®

İÇİNDEKİLER

BÖLÜM 1.....	5
---------------------	----------

Rasch Analizi: Teoriden Sağlık Araştırmalarına

Funda Çatan İnan

BÖLÜM 2.....	22
---------------------	-----------

**Comparison of Supervised Machine Learning Algorithms on Classification
of Skin Cancer Data**

Sinan Saraçlı



BÖLÜM 1

Rasch Analizi: Teoriden Sağlık Araştırmalarına

Funda Çatan İnan¹

1. GİRİŞ

Ölçme ve ölçüm araçları, toplumun her alanında önemli bir yer tutmaktadır. Nesnelerin ve canlıların ağırlığını belirlemek için kullanılan teraziler, kıyafetlerin ölçümünde kullanılan cetveller ve mezuralar, yemek yapımında gerekli olan malzeme miktarlarını ölçmek için kullanılan kaplar, vücut ve çevre sıcaklığını değerlendiren termometreler gündelik hayatta sıkça başvurulan araçlardır. Bu araçların ortak özelliği, doğru ölçüm yapılabilmesi için kalibrasyona, yani güvenilir standartlara dayanıyor olmalarıdır.

Buna karşın, doğrudan gözlemlenmesi mümkün olmayan değişkenlerin ölçümü daha karmaşık bir süreçtir. Psikososyal özellikler ya da yaşam kalitesi gibi bilişsel faktörler, doğrudan gözlenemeyen yapılar olup literatürde gizil değişkenler olarak adlandırılmaktadır. Gizil değişkenler, doğrudan ölçülememeleri nedeniyle nesnel ve güvenilir biçimde değerlendirilmesi en güç yapılardan biridir (Bond & Fox, 2021).

Bu noktada, klasik ölçme yaklaşımlarının sınırlılıkları ortaya çıkmaktadır. Doğrudan gözlenemeyen gizil yapıları daha geçerli ve güvenilir biçimde ölçmek amacıyla geliştirilen Madde Tepki Kuramı, gizil özellikleri niceliksel olarak ifade etmeyi amaçlayan matematiksel bir yaklaşımdır. Bu kuramın temel varsayımı, bir bireyin bir maddeye verdiği yanıtın, bireyin yetenek düzeyi ile maddenin özellikleri arasındaki farkın bir fonksiyonu olmasıdır.

Madde Tepki Kuramı kapsamında Rasch analizi, madde zorluğu ve kolaylığını temel parametre olarak ele almasıyla özgün bir konuma sahiptir. Danimarkalı matematikçi Georg Rasch tarafından 1960 yılında geliştirilen bu model, çalışılan veri setinin oluşturulan ölçüm modeline uygun olup olmadığını araştırır (Rasch, 1960). Rasch modeli yardımı ile modele iyi uyum gösteren veriler elde edilebilir (Andrich, 2004). Bu modeller, bireylerin test maddelerine verdikleri yanıtları matematiksel olarak modelleyerek hem bireylerin yetenek düzeylerini hem de maddelerin zorluk parametrelerini aynı ölçek üzerinde tahmin etme olanağı sunmaktadır (Jenkinson vd., 2011; Khine, 2020). Böylece gizil değişkenlerin ölçümü mümkün hale gelmekte ve elde edilen verilerden daha sağlam, karşılaştırılabilir ve genellenebilir sonuçlar elde edilmektedir. Bir değişkenin Rasch ölçümü yapılırken, maddeler o değişken için en kolaydan en

¹ Dr. Öğr. Üyesi, Bilecik Şeyh Edebali Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, ORCID: 0000-0003-3974-7577

zora doğru artan bir süreklilik üzerinde kavramsal olarak sıralanacak şekilde tasarlanır (Waugh, 2010).

Başlangıçta Rasch modeli eğitim alanında geliştirilmiş, daha sonra sosyal bilimlerde ve özellikle sağlık alanında hasta tarafından bildirilen sonuçların değerlendirilmesinde yaygın biçimde kullanılmaya başlanmıştır (Boone vd., 2014). Günümüzde ise Rasch modeli, klinik araştırmalar ve halk sağlığı çalışmaları kapsamında giderek daha fazla kullanılmaktadır. Fizik tedavi ve rehabilitasyon (Jimbo vd., 2025), yaşam kalitesi (Tennant vd., 2004; Lin vd., 2019; Kangwanrattanakul, 2025; Gu vd., 2025) ve farklı sağlık çıktıları gibi gizil değişkenlerin içsel yapı geçerliliğinin incelenmesinde geçerlik ve güvenirliği arttıran güçlü bir araç olarak kabul edilmektedir. Rasch analizi hem yeni ölçek geliştirme süreçlerinde hem de mevcut ölçeklerin geçerlik ve güvenirlik analizlerinde başvurulacak etkili bir yöntem olarak öne çıkmaktadır.

Bu kitap bölümünün amacı, eğitim, psikoloji ve sağlık alanında yaygın olarak kullanılan Rasch modeline bir giriş sunmaktır. Kitap bölümünde Rasch analizi teorik açıdan ele alınacak, model türleri ve analiz çıktıları değerlendirilecek, Rasch analizinin uygulanmasında kullanılacak yazılımlar tanıtılacak ve biyoistatistikteki uygulama alanları ortaya konulacaktır.

2. RASCH ANALİZİNİN AVANTAJLARI

Rasch modeli, bireylerin yetenek düzeyleri ile maddelerin güçlük düzeylerini ortak bir ölçüm ekseninde inceleyen bir analiz yöntemidir (Debelak vd., 2022). Model, olasılık kuramına dayalı olarak, bireyin belirli bir maddeye doğru cevap verme olasılığı, bireyin yetenek düzeyi ile madde güçlüğü arasındaki farka göre tahmin edilir. Bu fark logit adı verilen, negatiften pozitifte doğru uzanan sürekli bir ölçek üzerinde ifade edilir. Böylece bireyler yeteneklerine, maddeler ise zorluk düzeylerine göre aynı ölçekte sıralanır. Bir bireyin bir maddeye doğru yanıt verebilmesi, yetenek düzeyinin ilgili maddenin güçlük düzeyinden yüksek olmasına bağlıdır (Elhan ve Atakurt, 2005; Tennant ve Conaghan, 2007; Pallant ve Tennant, 2007).

Rasch modelinde yanıt yapısının Guttman tipi olması beklenir. Bu hiyerarşik ölçüm yaklaşımında, zor bir maddeye doğru yanıt veren bireyin, daha kolay maddelere de yüksek olasılıkla doğru yanıt vermesi öngörülür. Dolayısıyla, yüksek yetenekli bireylerin düşük yeteneklilere kıyasla daha zor maddeleri doğru cevaplama olasılıkları daha yüksektir (Elhan ve Atakurt, 2005; Tennant ve Conaghan, 2007).

Klasik Test Kuramı'nda toplam puan, maddelere verilen cevapların doğrudan toplanmasıyla elde edilir. Fakat bu yaklaşım bazı bazı sınırlılıklar içermektedir. Ölçek kategorilerinin eşit aralıklı olmaması, maddelerin farklı zorluk düzeylerine sahip olması, kayıp verilerin bulunması ve ham puanların doğrusal ölçeklerle ifade edilememesi bu yöntemin sınırlılıklara örnek olarak gösterilebilir. Rasch analizi, söz konusu sorunların büyük ölçüde aşılmasına olanak sağlayan güçlü bir yöntemidir. Ayrıca, analizlerin yalnızca gözlenen veriler üzerinden

gerçekleştirilebilmesi ve kayıp veriler için ek düzeltmelere ihtiyaç duyulmaması yöntemin önemli avantajları arasında yer almaktadır (Elhan ve Atakurt, 2005).

Klasik Test Kuramı ve Rasch analizi, yetenek kestirimi açısından farklılık göstermektedir. Likert tipi ölçeklerden elde edilen veriler, Klasik Test Kuramında eşit aralıklı ölçek gibi değerlendirilerek toplam puan üzerinden bireylerin yetenek düzeyleri belirlenir. Ancak verilerin sıralı yapısı nedeniyle bu yaklaşım yanlılıklara yol açabilir (Wang vd., 2025). Buna karşılık Rasch analizi, Likert ölçeklerini eşit aralıklı logit ölçeklerine dönüştürerek yetenek düzeylerini logit değerleri üzerinden tahmin etmektedir. Böylece iki yaklaşımda elde edilen dağılım özellikleri (çarpıklık, basıklık) farklılık göstermektedir. Ayrıca, Klasik Test Kuramında tüm maddeler yetenek kestirimine eşit katkıda bulunurken, Rasch analizinde bireylerin yetenek düzeyi madde parametrelerinden bağımsız olarak kestirilir.

Rasch modeli, araştırma bulgularını daha etkili bir şekilde görselleştirmeye olanak tanır. Wright haritaları kullanılarak bireylerin veya bir hasta grubunun yetenek düzeyi görsel olarak sunulabilir. Haritada bireylerin yetenek düzeyi ile maddelerin zorluk düzeyi aynı eksen üzerinde gösterilir. Bireyin ya da grubun ortalamasının altında kalan maddeler doğru yanıtlanması beklenen maddeleri, üstünde kalanlar ise doğru yanıtlanması beklenmeyen maddeleri temsil eder (Boone, 2020).

Bunun yanı sıra Rasch analizi, ölçüm yanlılığını araştırmak için de kullanılabilir. Örneğin, geliştirilmiş bir testin ya da ölçeğin erkek ve kadın katılımcıları aynı şekilde ölçüp ölçmediği incelenebilir. Bu tür karşılaştırmalarda testin her iki grup için de aynı işlevi yerine getirdiğinden emin olmak önemlidir. Rasch analizinde bu amaçla diferansiyel madde fonksiyonu (DIF) yöntemi uygulanır (Boone, 2020).

3. RASCH MODELİNİN VARSAYIMLARI

Rasch modelinin temel varsayımları aşağıda özetlenmiştir (Rasch, 1960; Pallant vd., 2006; Tennant vd., 2004):

- Yerel bağımsızlık: Aynı gizil özellik düzeyine sahip bireylerin, maddelere verdikleri yanıtlar birbirinden bağımsızdır. Gizil değişken kontrol altına alındıktan sonra, bireysel maddelere verilen yanıtlar arasında anlamlı bir ilişki bulunmamalıdır.
- Tek boyutluluk: Testte yer alan her bir madde yalnızca tek bir yeterliği ölçmekte olup, modelin tek boyutluluk koşulunu sağlaması gerekmektedir.
- Monotonluk: Bireylerin doğru yanıt verme olasılığı, yetenekleri arttıkça monoton olarak artar.
- Kişiden bağımsız madde kestirimleri: Madde konumları, bu kestirimlerin yapılmasında kullanılan bireylerden bağımsızdır.
- Maddeden bağımsız birey kestirimleri: Bireylerin konumları, bu kestirimlerin yapılmasında kullanılan maddelerden bağımsızdır.

Verilerin bu koşulları karşılaması, logit ölçeğinde madde ve kişi konumlarının anlamlı bir şekilde yorumlanmasını ve kullanılmasını destekler.

4. RASCH MODELLERİ

Rasch modellerinin çıkış noktasını, eğitim ölçmeleri alanında öğrencilerin başarı testlerindeki performansları oluşturmaktadır. Klasik test kuramında elde edilen toplam puanlar, çoğunlukla yalnızca gözlenen verilere bağlıdır ve bireylerin gerçek yeterlik düzeylerini tam anlamıyla yansıtmayabilir. Rasch modeli ise bireyin bir maddeye doğru cevap verme olasılığını, bireyin yetenek düzeyi ile maddenin zorluk parametresi arasındaki farka dayalı olarak tanımlar. Böylece, ölçümler madde bağımsızlığı ve kişi bağımsızlığı ilkeleri çerçevesinde daha genel geçer bir yapı kazanır.

İki ve daha fazla cevap kategorisi için farklı Rasch modelleri bulunmaktadır. Ölçekteki maddeler, iki sonuçlu ise ikili Rasch modeli (dichotomous model); ikiden fazla sonuçlu ise kısmi kredi modeli (partial credit model) ya da derecelendirme ölçeği modeli (rating scale model) kullanılmaktadır.

4.1. İkili Rasch Modeli

İkili Rasch modeli Rasch ailesinin en temel modelidir. İki kategorili maddeler (evet/hayır, katılıyorum/katılmıyorum, var/yok vb.) için kullanılan Rasch modeli, bireyin bir maddeyi doğru yanıtlama olasılığını, bireyin yetenek düzeyi (θ_i) ile maddenin zorluk parametresi (β_j) arasındaki farkın lojistik fonksiyonu olarak tanımlamaktadır. Bu model, Danimarkalı matematikçi Georg Rasch tarafından 1960 yılında geliştirilmiştir (Rasch, 1960).

Buna göre, birey tek basamağı başarıyla geçtiğinde “1” puan, başaramadığında ise “0” puan almaktadır. Bu modelde yalnızca maddenin zorluk parametresi (β_j) değil, aynı zamanda bireyin yetenek düzeyi (θ_i) de tahmin edilmektedir. Dolayısıyla, incelenen özelliği θ_i olan i. bireyin, zorluk düzeyi β_j olan j. maddeyi doğru yanıtlama ve buna bağlı olarak 1 puan alma olasılığı aşağıdaki formül ile ifade edilmektedir (4.1):

$$P(x_j = 1|\theta_i) = \frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} \quad (4.1)$$

4.2. Kısmi Kredi Modeli

Kısmi kredi modeli, ikiden fazla kategorili maddelere verilen yanıtların değerlendirilmesinde kullanılan bir Rasch modelidir (Wright & Masters, 1982). Bu modelde, yanıtlar yalnızca doğru/yanlış ikiliğine indirgenmez; bunun yerine doğruluk kavramı genişletilerek ara düzey kategoriler üzerinden derecelendirme yapılır. Böylece orta düzeyde yeteneğe sahip bireylerin de kısmen kredilendirilmesi mümkün hale gelmektedir.

Kısmi kredi modeli özellikle aşağıdaki durumlarda yaygın olarak tercih edilir:

- Çok aşamalı çözümleme gerektiren maddelerde,

- Farklı aşamaların doğru cevaplanma derecelerinin önemli olduğu durumlarda,
- Likert tipi ölçeklerde yanıt kategorileri arasındaki aralıkların maddeden maddeye farklılık gösterdiği ölçümlerde.

Bu model kapsamında, yetenek düzeyi θ olan i. bireyin, j. maddeden x yanıtını alma olasılığı aşağıdaki formül ile ifade edilmektedir (4.2):

$$P_{ijx} = \frac{\exp \sum_{k=0}^x (\theta_i - \beta_{jk})}{\sum_{k=0}^{m_i} \exp \sum_{t=0}^k (\theta_i - \beta_{jt})}, \quad x = 0, 1, \dots, m_i \quad (4.2)$$

Denklem 4.2’de

θ_i : i.bireyin yetenek düzeyi,

β_j : j.maddenin zorluk parametresi,

m_i : maddenin maksimum kategori sayısıdır.

4.3. Derecelendirme Ölçeği Modeli

Derecelendirme ölçeği modeli, özellikle Likert tipi ölçeklerde sıklıkla kullanılmaktadır. Bu modelde, elde edilen veriler sıralama ölçeği düzeyinde olup eşit aralıklı ölçek düzeyine dönüştürülmesi gerekmektedir. Derecelendirme ölçeği modeli ile sıralamalı ölçeği düzeyindeki veriler logaritmik dönüşüm ile eşit aralıklı logit ölçeğine dönüştürülür.

5’li Likert tipinde (Kesinlikle katılmıyorum=1, Katılmıyorum=2, Kararsızım=3, Katılıyorum=4, Kesinlikle katılıyorum=5) bir ölçeği ele alalım. Bir birey ‘Katılmıyorum’ kategorisini seçtiğinde ‘Kesinlikle Katılmıyorum’ yerine “Katılmıyorum” kategorisini seçerek ilk basamağı tamamlamış, ancak ‘Katılmıyorum’ yerine ‘Kararsızım’ kategorisini seçemediği için ikinci basamakta başarısız olmuştur.

Derecelendirme ölçeği modelinde, basamakların zorluk düzeylerinin tüm maddeler için aynı olduğu varsayılmaktadır. Bu varsayım, Kısmi kredi modeli içinde madde basamaklarının aşağıdaki şekilde ayrıştırılmasıyla elde edilir (4.3):

$$\beta_{jk} = \beta_j + \tau_k \quad (4.3)$$

Bu eşitlik, kısmi kredi modeli denklemi içerisine yerleştirildiğinde derecelendirme ölçeği modeli elde edilir. Bu modele göre, i. bireyin j. maddeden x kategorisini seçme olasılığı şu şekilde formüle edilir (4.4):

$$P_{ijx} = \frac{\exp \sum_{k=0}^x (\theta_i - (\beta_j + \tau_k))}{\sum_{k=0}^{m_i} \exp \sum_{t=0}^k (\theta_i - (\beta_j + \tau_t))}, \quad x = 0, 1, \dots, m_i \quad (4.4)$$

Burada:

θ_i : i.bireyin yetenek düzeyi,

β_j : j.maddenin zorluk parametresi,

m_j : maddenin maksimum kategori sayısı,

τ_k : k.kategori eşik parametresidir (Bond & Fox, 2021).

5. RASCH ANALİZİNDE ÇIKTILAR VE YORUMLAMA

Rasch analizinin uygulanmasında sıklıkla kullanılan Winsteps programı ile oluşturulan Wright haritalarında, logit birimleri ile ölçeklendirilmiş dikey bir çizgi bulunur. Logit ölçeği aralıklı bir ölçme düzeyine sahip olduğundan, dikey ölçeğin herhangi bir noktasındaki eşit uzaklıklar aynı büyüklüğe karşılık gelmektedir (Bond & Fox, 2021).

Rasch analizi yazılımlarının birçok sürümünün ürettiği haritalarda maddeler madde numaraları ile, her bir bireyin performansı ise bir 'X' sembolü ile gösterilmektedir. Küçük veri setlerinde 'X' tek kişiyi gösterirken, büyük veri setlerinde birden fazla kişiyi gösterebilir; X'in temsil ettiği kişi sayısı da '#' karakteri ile haritanın altında yazılır.

Her madde ve birey, kestirilen değerine göre logit ölçeği üzerinde konumlandırılır. Daha yüksek değere sahip bireyler daha yetenekli, daha yüksek değere sahip maddeler ise daha zordur (Bond & Fox, 2021; Boone, 2014; Boone, 2016). Örneğin; eğitim testlerinde kişi parametresi öğrencinin yeteneğini, madde parametresi ise maddenin kolaylığını temsil eder; yetenek düzeyi yükseldikçe ve madde daha kolay oldukça, doğru yanıt verme olasılığı artar. Sağlık bilimlerinde ise kişi parametresi bireyin depresyon düzeyini, madde parametreleri ise depresyonla ilişkili belirli semptomları yaşama riskini gösterebilir.

Her Wright haritasında, dikey çizginin her iki yanında 'M, S ve T' sembolleri bulunur. Haritanın madde tarafındaki 'M' maddelerin güçlük ortalamasını gösterir. 'S' harfi ortalamanın bir standart sapma üstü ve altını, 'T' harfi ise ortalamanın iki standart sapma üstü ve altını temsil eder. Aynı harfler kişi ölçümleri için de benzer bilgiyi sunmak amacıyla kullanılır (Boone vd., 2014).

MEASURE

PERSON - MAP - ITEM

<more>|<rare>

2

+

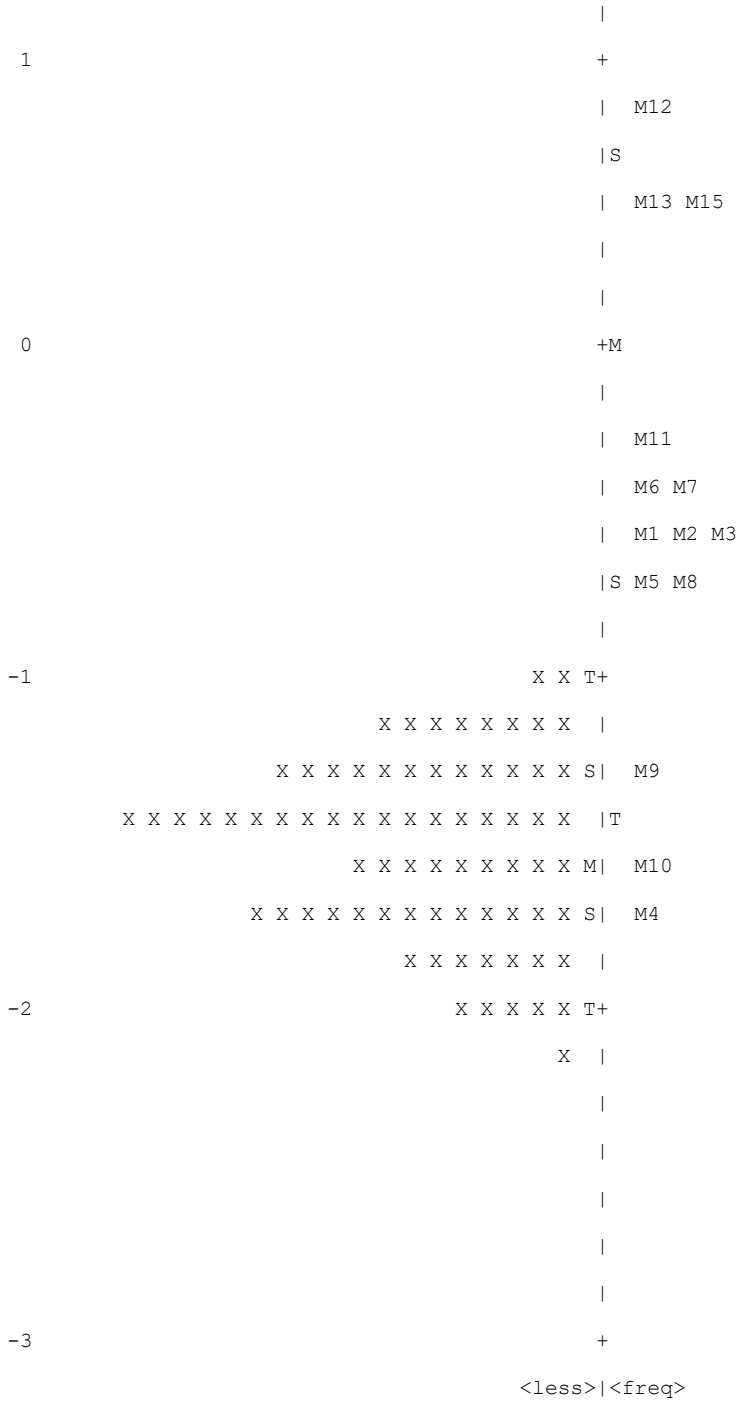
|

|

|

|T

|



Figür 1. Depresyon ölçeği için Wright haritası

Bu çalışmada hastaların depresyon düzeylerini ölçmeye yönelik 15 maddelik bir ölçek 75 kişilik bir örnekleme Winsteps programının ücretsiz versiyonu olan Ministeps'te uygulanmıştır. Çalışmada kullanılan veriler ise Excel'de simülasyon yolu ile elde edilmiştir. Rasch analizi sonucunda elde edilen Wright haritası Figür 1'de sunulmuştur. Haritanın sol tarafında bireylerin depresyon düzeyleri, sağ tarafında ise ölçek maddelerinin zorluk parametreleri yer almaktadır. Katılımcıların büyük çoğunluğu -2 ile -1 logit aralığında yoğunlaşmaktadır; bu da örneklemin genel olarak düşük düzeyde depresif semptomlar bildirdiğini göstermektedir. Maddelerin ise çoğunlukla 0 ile +1 logit aralığında yer aldığı (ör. madde 6, 7, 11, 12), yalnızca birkaç maddenin görece kolay (ör. madde 9, 10) veya zor (ör. madde 13, 15) olduğu dikkat çekmektedir. Harita, ölçeğin örneklem için görece zorlayıcı olduğunu göstermektedir. Özellikle düşük semptom düzeylerini ayırt etmede sınırlı hassasiyete sahiptir.

Rasch uyum istatistikleri, bireylerin ve verdikleri yanıtların Rasch ölçme modelinin öngördüğü yanıt düzeniyle ne kadar örtüştüğünü gösterir. Madde uyum istatistikleri, bir maddenin ait olduğu modele uygunluğu uyum içi ortalama kare (infit MnSq) ve uyum dışı ortalama kare (outfit MnSq) değerleri ile kontrol edilmektedir. Uyum içi ve uyum dışı istatistikleri, sıfır ile sonsuz arasında değişen ki-kare değerleri olarak hesaplanır. Rasch modeline tam uyum gösteren maddelerin ortalama kare değerleri 1'dir (Linacre, 2002). Literatürde, uyum içi ve uyum dışı ortalama kare değerlerinin 0,5–1,5 aralığında; DIF karşılığının 0,5'in altında; madde ve kişi güvenirliğinin 0,70'in üzerinde; ayırma indekslerinin ise 2'nin üzerinde olması gerektiği belirtilmektedir (Lin vd., 2019a; 2019b; Mamun vd., 2022). Bununla birlikte, bazı çalışmalarda uyum içi ve uyum dışı ortalama kare değerleri için kabul edilebilir aralık 0,6-1,4 olarak önerilmektedir (Bond & Fox, 2021). Bir maddenin ortalama kare değerinin bu değerlerin üzerinde olması, maddenin uyumsuz (underfit) olduğunu göstermektedir. Bunun yanı sıra, ortalama kare değerinin düşük çıkması ise maddenin aşırı uyumlu (overfit) olduğunu ifade etmektedir (Wright, 1994; Linacre, 2002; Bond & Fox, 2021).

Maddelerin kolay/zorluk durumlarına ait standart hata (RMSE), aşırı uçlarda bulunan değerler dışındaki tüm değerler için ölçme hatası olarak bulunur. Rasch analizi sonucunda elde edilen uyum istatistiklerinde ayrıca uyum içi ve uyum dışı ZSTD değerleri incelenmelidir. Bu değer, madde ve kişiler için standart ağırlıklandırılmamış uyum istatistiği olarak tanımlanır. Değer aralığı (-2, 2) olan ZSTD, en iyi uyumu ortalamanın 0, standart sapmanın 1 olduğu durumda sağlar.

Rasch modelinde güvenirlik hem bireyler hem de maddeler için kestirilmektedir. Kişi güvenirliği, ölçeğin bireyleri ne kadar iyi ayırt edebildiğini gösteren bir kestirimdir (Wright & Masters, 1982). Bu kavram, sıfır ile sonsuz arasında değer alan ve kişi ölçümlerinin standart hata birimleri cinsinden dağılımını ifade eden kişi ayırma indeksi ile ifade edilir (Panayides vd., 2013). Kişi ayırma indeksi kişileri yeteneklerine göre sınıflandırır. Madde güvenirliği ise, maddelerin yanıtlayıcı örneklem tarafından ne ölçüde ayırt edilebildiğini gösterir. Madde ayırma indeksi ise madde hiyerarşisini belirler (Linacre, 2009). Kişiler için güvenirlik katsayısı kişilerin sıralanabilirliğinin yeniden

üretilebilirliğini, maddeler için güvenilirlik katsayısı maddelerin zorluk düzeylerinin yeniden üretilebilirliğini inceler (Lin vd., 2019). Dolayısıyla etkili bir ölçüm için hem yüksek madde ayırt ediciliği hem de yeterli kişi ayırma güvenilirliği gereklidir (Wright & Masters, 1982).

Sonuç olarak, Rasch analizi çıktılarına ilişkin Wright haritaları, uyum istatistikleri ve güvenilirlik ölçütleri birlikte değerlendirildiğinde, ölçeğin hem bireyler hem de maddeler açısından ölçüm kalitesi hakkında kapsamlı bilgi sunmaktadır. Bu göstergeler, sağlık alanındaki ölçeklerin geçerlik ve güvenilirliğinin sağlanmasında kritik rol oynamaktadır.

6. KULLANILAN YAZILIMLAR

Gizil değişkenlerin ölçülmesi ve analiz edilmesi, klinik araştırmalar ve epidemiyoloji alanında giderek önem kazanmaktadır. Özellikle hasta bildirimine dayalı sonuçların kullanımının artmasıyla birlikte, Rasch analizi sağlık bilimlerinde yaygın biçimde kullanılmaya başlanmıştır. Geleneksel istatistik paket programları, çoğu zaman bu özel alanda madde analizi yapmaya veya model parametrelerini tahmin etmeye olanak tanımadığı için araştırmacılar, bir ölçeğin ölçme özelliklerini analiz edebilmek için çeşitli özel yazılımlara ihtiyaç duymaktadır. Rasch analizini uygulayabilmek için yaygın kullanılan programlar özetlenmiştir:

ConQuest: Wu ve arkadaşları (2007) tarafından geliştirilen ConQuest, Quest yazılımının geliştirilmiş sürümüdür (Wu vd., 2007). İkili, derecelendirme ölçeği, kısmi kredi ve çok boyutlu Rasch modellerinin yanı sıra, ConQuest programı çok değişkenli Rasch modellerini de analiz edebilir. Özellikle eğitim ölçmelerinde büyük ölçekli veri setlerinin analizi için tercih edilmektedir. Madde ve kişi parametrelerini aynı anda tahmin edebilmesi ve çok boyutlu yapıları modelleyebilmesi en büyük avantajlarından.

TAM: Açık kaynak kodlu R ortamında kullanılan TAM paketi, birleşik ve marjinal maksimum olasılık yöntemleri ile Rasch modellerini tahmin etmektedir. Çok boyutlu Rasch modelleri ve karma modeller üzerinde çalışılabilir. R kullanıcıları için ücretsiz ve iyi bir alternatiftir.

Winsteps: WINSTEPS, Rasch analizi için en yaygın kullanılan programlardan biridir. Chicago Üniversitesi'nde geliştirilmiş olan programın son güncellemesi Linacre tarafından 2009 yılında yapılmıştır (Linacre, 2009) Tek boyutlu Rasch modellerini çalıştırabilir; hem ikili (dichotomous) hem de çok kategorili (polytomous) veri setlerini analiz edilmektedir. Parametrelerin tahmini için birleşik maksimum olasılık yöntemini kullanır. Program, madde ve kişi parametrelerinin tahmin edilmesinin yanı sıra Wright haritası (madde-kişi haritası) oluşturma ve model uyum istatistiklerini raporlama gibi gelişmiş özellikler sunmaktadır. Ayrıca, sınırlı işlevselliğe sahip ücretsiz öğrenci sürümü de mevcuttur.

Digram: Kreiner (2003) tarafından geliştirilen Digram, ikili ve çoktan seçmeli maddeler için Rasch analizi yapabilen ücretsiz bir yazılımdır. Kullanımı nispeten basit olup temel analiz ihtiyaçlarını karşılamaktadır.

Facets: Linacre tarafından geliştirilen ve çok-yüzlü Rasch ölçme modeli) uygulamaları için tasarlanmıştır (Linacre, 1994). Klasik Rasch modellerinden farklı olarak yalnızca maddeleri değil; aday, değerlendirici, görev ve ortam gibi ek boyutları da modele dahil eder. Özellikle performansa dayalı değerlendirmelerde ve öznel yargının söz konusu olduğu durumlarda tercih edilmektedir. Facets, madde ve kişi haritaları, değerlendirici yanlılık analizleri ve model uyum istatistikleri gibi ayrıntılı çıktılar sunar.

RUMM2030: RUMM2030 özellikle sağlık bilimleri alanındaki araştırmacılar tarafından sıklıkla tercih edilmektedir. Grafiksel ara yüzü ile kullanıcı dostudur. Parametre tahmini, eşleşmeli koşullu maksimum olasılık yöntemine dayanmaktadır. Program ile ikili Rasch modeli, derecelendirme ölçeği modeli ve kısmi kredi modeli dahil olmak üzere çeşitli Rasch modeline uygulanabilmektedir. Program, madde uyum istatistiklerini detaylı biçimde raporlamakta ve grafiksel çıktılar sağlamaktadır. Ayrıca tek boyutluluk varsayımının test edilmesi ve yanlılık (DIF) analizleri için de işlevsel modüller içermektedir (Andrich vd., 2010).

7. BİYOİSTATİSTİKTE UYGULAMA ALANLARI

Biyoistatistik, sağlık bilimlerinde yalnızca klinik araştırmaların değil, ölçme, değerlendirme ve karar verme süreçlerinin de bilimsel bir temele dayandırılmasını amaçlayan bir disiplindir. Günümüzde özellikle hasta bildirimine dayalı sonuç ölçümleri, yaşam kalitesi ölçekleri, psikososyal değerlendirme araçları ve klinik ölçeklerin geliştirilmesi gibi alanlarda ölçüm hatalarını en aza indiren, güvenilir ve geçerli ölçme yöntemlerine duyulan ihtiyaç artmaktadır. Bu açıdan bakıldığında, em yeni ölçek geliştirme süreçlerinde hem de mevcut ölçeklerin psikometrik özelliklerinin incelenmesinde önemli avantajlar sunmaktadır (Elhan Atakurt, 2005; Pallant & Tennant, 2007; Tennant & Conaghan, 2007; Jenkinson vd., 2011). Özellikle sıralı yanıt kategorilerinin aralıklı ölçeklere dönüştürülmesinde Rasch modeli klasik test kuramına dayalı yaklaşımlara güçlü bir alternatif oluşturmaktadır.

Rasch analizinin biyoistatistikte kullanım amaçlarından bazıları aşağıdaki gibi sıralanabilir:

- **Gizil değişkenlerin ölçülmesi**

Rasch modeli, gözlenemeyen özelliklerin (örneğin depresyon düzeyi, ağrı şiddeti, işlevsel bağımsızlık gibi) doğrudan ölçülemeyen ancak maddelere verilen tepkiler aracılığıyla tahmin edilen boyutlarını ölçmeye olanak tanır.

- **Ordinal verilerin aralıklı ölçeklere dönüştürülmesi**

Klinik araştırmalarda sıklıkla kullanılan Likert tipi ölçekler ordinal yapıya sahiptir. Rasch modeli sayesinde bu tür veriler, logit ölçeğinde aralıklı ölçümlere dönüştürülerek daha hassas analiz yapılabilir

- **Ölçek geliştirme ve uyarlama**

Ölçek maddelerinin zorluk düzeylerinin, işlevselliğinin ve model uyumunun değerlendirilmesi, tek boyutluluk ve ölçme eşdeğerliği gibi psikometrik özelliklerin test edilmesi Rasch analiziyle mümkün hale gelir.

- **Klinik karar süreçlerine katkısı**

Tedavi etkinliğinin değerlendirilmesinde kullanılan ölçeklerin ölçme hassasiyeti, hasta grupları arasında farkların doğru şekilde ortaya konulması için kritik öneme sahiptir.

Rasch analizi, veri setlerinde uyum istatistikleri (infit, outfit) ile madde performansının incelenmesinde, diferansiyel madde fonksiyonu ile maddelerin farklı alt gruplarda eşit görev görüp görmediğinin test edilmesi, kişi madde-haritası ile bireylerin yetenek düzeyleri ile maddelerin zorluk düzeylerinin aynı ölçek üzerinde karşılaştırılması, güvenilirlik katsayıları ile (ayırma indeksi, ayırma güvenilirliği) ile ölçeğin ayırt edici gücünün değerlendirilmesi alanlarında kullanılmaktadır.

Sağlık bilimlerinde Rasch analizinin uygulama alanlarına; yaşam kalitesi ölçekleri (Tennant vd., 2004; Lin vd., 2019; Kangwanrattanakul, 2025; Gu vd., 2025), anksiyete, depresyon (Pallant vd., 2006; Faulkner vd., 2025), tükenmişlik sendromu (Bartholomew vd., 2025) ve uyku bozukluğu (Mamun vd., 2021; Westergren vd., 2025) gibi psikiyatrik hastalıkların değerlendirilmesinde kullanılan ölçekler, fizik tedavi ve rehabilitasyon (Jimbo vd., 2025) alanında uygulanan yöntemlerin etkinliğinin ölçümü, epidemiyolojik çalışmalarda toplumsal sağlık göstergeleri, klinik denemelerde hasta bildirimine dayalı sonuçların ölçümü örnek olarak verilebilir.

8. SONUÇ VE GELECEK YÖNELİMLER

Rasch analizi, biyoistatistikte ölçme kalitesini artıran, klinik kararların bilimsel doğruluğunu güçlendiren ve sağlık araştırmalarına metodolojik derinlik kazandıran bir yöntemdir. Sosyal bilimlerden sağlık araştırmalarına kadar pek çok alanda ölçme araçlarının geliştirilme ve kullanıma biçimini köklü biçimde etkilemiştir. Özellikle hasta odaklı sonuçların giderek önem kazandığı yeni tıp araştırmalarında, Rasch modelinin kullanımı sadece metodolojik bir tercih olarak değil, aynı zamanda klinik geçerlilik açısından kritik bir rol oynayacaktır. Böylece Rasch analizi, hem ölçme bilimleri hem de uygulamalı sağlık araştırmaları için stratejik bir araç olarak konumunu güçlendirmektedir.

Gelecekte, yapay zekâ ve makine öğrenmesi tabanlı algoritmalar ile Rasch analizi sonuçlarının bütünleştirilmesi sağlık değerlendirmeleri için yeni fırsatlar sunabilir. Bu sayede bireyselleştirilmiş sağlık değerlendirmeleri, eğitim planlamaları ve tedavi stratejileri daha doğru ve hassas biçimde oluşturulabilecektir. Aynı zamanda, çok boyutlu Rasch modelleri, latent regresyon ve hiyerarşik Bayes yöntemleriyle entegrasyon, birey-madde etkileşimlerinin daha karmaşık yapılarının incelenmesine olanak tanıyabilir. Son

olarak, Rasch analizi hakkındaki farkındalığın artırılması, yöntemle ilgili yanlış uygulamaların önlenmesini ve bilimsel güvenilirliğin yükselmesini sağlayacaktır.

Referanslar

1. Andrich D., Sheridan B., Luo G. (2010). RUMM2030, Computer software and manual, RUMM Laboratory, Perth, Australia.
2. Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care*, 42(1), 7–16.
3. Bartholomew, E. J., Medvedev, O. N., Petrie, K. J., & Chalder, T. (2025). Rasch analysis of the hospital anxiety and depression scale in patients with chronic fatigue syndrome. *Journal of Psychosomatic Research*, 112370.
4. Bond, T., & Fox, C. (2021). *Applying the rasch model: Fundamental measurement in the human sciences.4th Edition*. Mahwah, NJ:Lawrence Erlbaum Associates.
5. Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how?. *CBE—Life Sciences Education*, 15(4), rm4.
6. Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch analysis in the human sciences. Dordrecht: Springer.
7. Boone, W.J. (2020). Rasch Measurement Applications in Quantitative Educational Research. Khine, M.S (Ed.) Singapore: Springer
8. Debelak, R., Strobl, C., Zeigenfuse, M.D. (2022). An Introduction to the Rasch Model with Examples in R. USA: Chapman & Hall/CRC.
9. Elhan, A. H., & Atakurt, Y. (2005). Ölçeklerin değerlendirilmesinde niçin Rasch analizi kullanılmalıdır?. *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, 58(01), 47-50.
10. Faulkner, J. W., Snell, D. L., & Siegert, R. J. (2025). Rasch analysis of the depression anxiety stress scales-21 (DASS-21) in a mild traumatic brain injury sample. *Brain injury*, 39(2), 136-144.
11. Gu, R., Zhu, L., Kong, J., Zhang, L., Song, M., Cheng, X., ... & Wang, H. (2025). Development of the Short Form for Chronic Hepatitis B Quality of Life Instrument (CHBQOL-SF) Using Delphi Method and Rasch Analysis. *Assessment*, 10731911251321922.
12. Jenkinson, C., Peters, M., & Bromberg, M. B. (Eds.). (2011). *Quality of life measurement in neurodegenerative and related conditions*. pp. 147-164. Cambridge University Press.
13. Jimbo, K., Miyata, K., Yuine, H., Takahama, K., Yoshimura, T., Shiba, H., ... & Shiraishi, H. (2025). Verification of item difficulty in the Capabilities of Upper Extremity Test using Rasch analysis for individuals with cervical spinal cord injury. *The Journal of Spinal Cord Medicine*, 1-9.

14. Kangwanrattanakul, K. (2025). Validation of the Thai World Health Organization Quality of Life-OLD (WHOQOL-OLD) among Thai older adults: Rasch analysis. *Scientific Reports*, 15(1), 12978.
15. Khine, M.S. (2020). Rasch Measurement Applications in Quantitative Educational Research. Khine, M.S (Ed.) Singapore: Springer
16. Kreiner S. (2003). Introduction to DIGRAM, Research report 03/10, Department of Biostatistics, University of Copenhagen, Copenhagen.
17. Lin, C. Y., Hwang, J. S., Wang, W. C., Lai, W. W., Su, W. C., Wu, T. Y., Yao, G. & Wang, J. D. (2019a). Psychometric evaluation of the WHOQOL-BREF, Taiwan version, across five kinds of Taiwanese cancer survivors: Rasch analysis and confirmatory factor analysis. *Journal of the Formosan Medical Association*, 118(1), 215-222.
18. Lin, C. Y., Imani, V., Broström, A., Nilsen, P., Fung, X. C., Griffiths, M. D., & Pakpour, A. H. (2019b). Smartphone application-based addiction among Iranian adolescents: A psychometric study. *International Journal of Mental Health and Addiction*, 17(4), 765-780.
19. Linacre J.M. (2009). Winsteps (version 3.68), Computer Software and manual, Winsteps.com, Beaverton, OR.
20. Linacre, J. M. (1994). Constructing measurement with a many-facet Rasch model. *Objective measurement: Theory into practice*, 2, 129-144.
21. Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch measurement transactions*, 16(2), 878.
22. Mamun, M. A., Alimoradi, Z., Gozal, D., Manzar, M. D., Broström, A., Lin, C. Y., ... & Pakpour, A. H. (2021). Validating insomnia severity index (ISI) in a Bangladeshi population: using classical test theory and Rasch analysis. *International journal of environmental research and public health*, 19(1), 225.
23. Pallant, J. F., Miller, R. L., & Tennant, A. (2006). Evaluation of the Edinburgh post natal depression scale using Rasch analysis. *Bmc Psychiatry*, 6(1), 28.
24. Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *British journal of clinical psychology*, 46(1), 1-18.
25. Panayides, P., Gavrielides, M., Galatopoulos, C., & Gavriliidou, M. (2013). Using Rasch measurement to create a quality of sleep scale for a non-clinical sample based on the Pittsburgh Sleep Quality Index (PSQI). *Europe's Journal of Psychology*, 9(1), 113-135.
26. Rasch, G. (1960). Probabilistic model for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research.

27. Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in health*, 7, S22-S26.
28. Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?. *Arthritis Care & Research*, 57(8), 1358-1362.
29. Wang, X., Cao, C. H., Liao, X. L., Jiang, X. Y., Griffiths, M. D., Chen, I. H., ... & Malas, O. (2025). Comparing the psychometric evidence of the Depression, Anxiety, and Stress Scale-21 (DASS-21) between Spanish and Chinese primary schoolteachers: insights from classical test theory and rasch analysis. *BMC psychology*, 13(1), 1-18.
30. Waugh, R. (2010). *Applications of Rasch measurement in education*. Nova Science Publishers.
31. Westergren, A., Hedin, G., Hagell, P., Bäckman, C., & Norell, A. (2025). A Rasch model based psychometric evaluation of different versions of the Minimal Insomnia Symptom Scale (MISS) among adults. *Sleep Science and Practice*, 9(1), 12.
32. Wright BD (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions* ;8:370.
33. Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
34. Wu M.L., Adams R.J., Wilson M.R., Haldane S. (2007). ConQuest (Version 2.0), Computer Software and manual, ACER, Camberwell, Australia.



BÖLÜM 2

Comparison of Supervised Machine Learning Algorithms on Classification of Skin Cancer Data

Sinan Saraçlı¹

IntroductionIn recent years, the rapidly advancing field of artificial intelligence algorithms has simplified the application of analytical methods used in scientific research to any existing dataset. While a high level of programming knowledge was previously required to understand deep learning algorithms, technological advancements have significantly facilitated the use of software. Consequently, deep learning and machine learning methods have started to be used in various medical disciplines, which has, in turn, enhanced the quality of the research conducted.

Machine Learning describes a set of methods commonly used to solve a variety of real-world problems with the help of computer systems, which can learn to solve a problem instead of being explicitly programmed to do so (Koza et al., 1996). For instance, instead of explicitly telling a computer system which words within an tweet would indicate it to contain a customer need, the system (given a sufficient set of training samples) learns the typical patterns of words and their combination which results in a need classification (Kühl et al., 2020).

Supervised Learning (SL), Unsupervised Learning (UL), Reinforcement Learning (RL) and Recommender Systems (RS) are the types of Machine learning. *SL* is based on the comparison of computed and expected output, that is learning refers to computing the error and adjusting the error for achieving the expected output. *UL* is termed as learned by its own by discovering and adopting, based on the input pattern. In this learning the data are divided into different clusters and hence the learning is called a clustering algorithm. *RL* is based on output with how an agent ought to take actions in an environment so as to maximize some notion of long-term reward. A reward is given for correct output and a penalty for wrong output and *RS* can be defined as a learning technique by virtue of which online user can customize their sites to meet customer's tastes (Das et. al., 2015).

Image classification which belongs to the main research content of image processing has a broad application prospect in many sciences, such as object recognition, content understanding and image matching (Jiao, et al., 2021). The

¹ Prof. Dr., Balıkesir University, ORCID: 0000-0003-4662-8031

Random Forest (RF), Extreme Gradient Boosting (XGB), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) are Machine Learning classifiers suitable for analyzing high-dimensional and complex data and developing valuable predictions. (Moldovanu, et al., 2025).

Dermatology, as a visually intensive field, is at the precipice of an AI revolution. The association for the advancement of AI defines it as “the scientific knowledge of the mechanisms underlying mind and intelligent behavior and its implementation in machines” (Hernández-Orallo et al.,2016). Because skin disease diagnosis is mostly based on visual perception, computer vision algorithms may be able to recognize skin lesions based on their morphology (Chan et al., 2020; Das et al., 2021). It is further predicted to surpass heart disease as the main cause of mortality and the biggest obstacle to extending life expectancy in the next decades. (Han et al.,2006; Minko et al, 2013; Siegel et al., 2023; Tiwari et al, 2023; Hasan et al, 2023a; Hasan et al. 2023b).

Evaluating both benign and malignant neoplasms is vital, particularly due to the increasing incidence of skin cancers and complications associated with the condition (Rogers et al., 2012; Glazer et al., 2017). Skin cancer has multiple etiologic factors, with UV radiation-induced DNA damage and oncogenesis being the most prominent. Preventive strategies, such as sunscreen application, are essential in reducing risk (Haddad et al., 2025). After evaluation by a trained healthcare provider, a biopsy is typically performed, and histopathological evaluation helps confirm the diagnosis and differentiate the type of skin cancer from other conditions (Elston et al 2016). Different types of skin cancer are given visually in Figure 1.

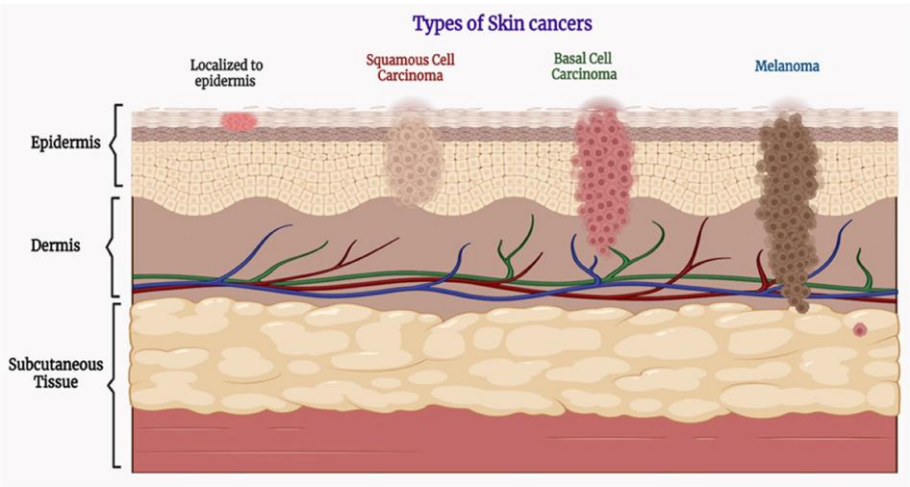


Fig.1. Different types of skin cancer (Source: Hasan et al. 2023b)

Materials and Methods

The skin cancer images used in this study were obtained from the Kaggle platform (<https://www.kaggle.com/>). The dataset is divided into two classes: benign and malignant. The dataset used in the study was split into training and test subsets, with an 80% and 20% ratio, respectively. Both classes were organized to be balanced within the training and test subsets.

Sample images from the dataset are presented in Figure 2. Benign samples are located in the first row, and malignant samples are in the second row. Examining the structure of the visuals, it is noteworthy that benign lesions have more regular boundaries and homogeneous colors, while malignant lesions exhibit more irregular borders and heterogeneous color distributions.

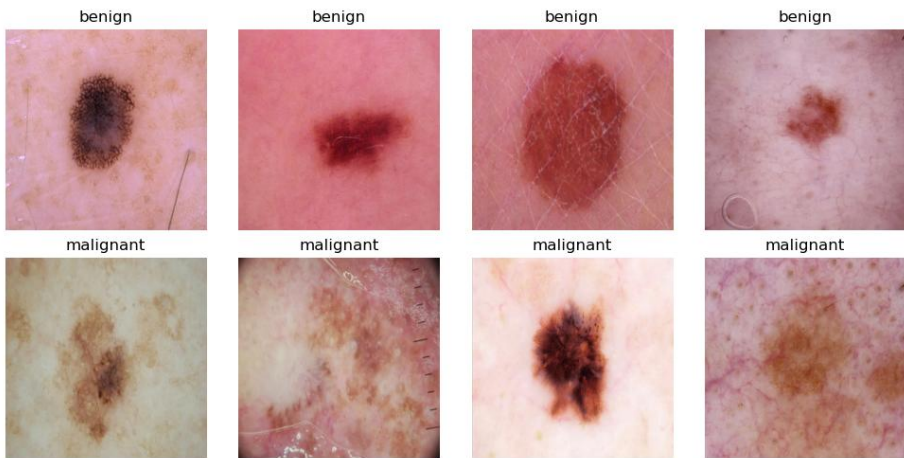


Figure 2. Sample images from the dataset

In the analysis section of the study, for Image Preprocessing, all images were rescaled to 224×224 pixel dimensions to ensure model compatibility. Furthermore, pixel values were normalized to the range of 0–1 during the training and testing phases. The Keras ImageDataGenerator class was used for loading and scaling the images.

For Feature Extraction, classification was not performed directly on the raw pixels; instead, color-based features were extracted to create a more meaningful feature space. Feature extraction was carried out using the following steps:

1. **Color Space Transformation:** Images were converted from the RGB color space to the CIE-LAB color space. This transformation allows for a color representation that is more compatible with human visual perception.
2. **Statistical Moments:** The mean, standard deviation, skewness, and kurtosis values were calculated for the L, A, and B channels of each

image. This resulted in a total of 12 statistical features (four statistical moments for each channel).

3. Histogram-Based Features: Normalized histograms with 256 bins were generated for the L channel in the range $[0,100]$ and for the A and B channels in the range $[-128,128]$. These histogram values quantitatively represented the color distribution of the images.
4. Feature Vector: The statistical moments and histogram-based features were combined for each image to obtain the final feature vector.

Classification Models

In this study, three different machine learning methods were utilized for the classification of the color-based features extracted from the skin cancer images: *Extreme Gradient Boosting* (XGBoost), *Logistic Regression*, and *Random Forest*. These models were selected due to their widespread use in the literature for both medical image analysis and general classification problems, their tendency to demonstrate high performance, and the different methodological advantages they offer.

XGBoost is an optimized version of the gradient boosting algorithm based on decision trees. Chen and Guestrin (2016) demonstrated that XGBoost is effective on sparse datasets, resistant to overfitting through regularization (L1 and L2), and computationally fast on large-scale datasets. The literature emphasizes that XGBoost offers strong generalization performance, particularly in cases with high-dimensional feature spaces (Nielsen, 2016). However, the model's primary disadvantages are the requirement for tuning a large number of hyperparameters and its limited interpretability (Zhang et al., 2020). In this study, XGBoost was chosen due to its ability to model the non-linear relationships within the high-dimensional feature set, which consists of color histograms and statistical moments.

Logistic Regression is one of the most frequently used linear methods for binary classification problems. It is widely applied in medicine and social sciences due to its simplicity, interpretability, and its ability to produce a probability output (Hosmer et al., 2013). Among the advantages of Logistic Regression are the direct interpretability of the model coefficients and the ability to adjust the decision threshold according to clinical applications. However, it may show limited performance on complex and non-linear data structures because of its assumption of a linear decision boundary (Peng et al., 2002). In this study, Logistic Regression was used as a fundamental baseline model for comparison with more complex models.

The *Random Forest* algorithm, proposed by Breiman (2001), is an ensemble learning method based on building numerous decision trees using random subsets of features and samples, and then combining their predictions through voting.

The most significant advantages of Random Forest are its high generalization success, its ability to model complex interactions between features, and its insensitivity to feature scaling (Cutler et al., 2007). Furthermore, it has been reported in the literature to show high performance, particularly in biomedical image classification (Rodriguez-Galiano, et al., 2012). However, it is noted that there may be bias in variable importance measures and that the model's interpretability is limited (Strobl et al., 2007). In this study, Random Forest was chosen because of its potential to capture the complex distributions of the histogram-based features.

Findings

Findings of the study is given below in Table 1 – Table 5 detailly. To compare the classification performances of the model, the Evaluation metrics are calculated as given below.

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Here, TP denotes true positive, TN true negative, FP false positive, and FN false negative.

Table 1. Image numbers in training and test subsets

Data set	Benign	Malignant	Total
Training	1440	1197	2637
Test	360	300	660

Table 1 indicates that, a total of 2,637 images were used for training and 660 images for testing in the dataset. It can also be seen that the class distributions are relatively balanced.

Table 2. Classification performance of XGBoost

Class	Precision	Recall	F1-Score
Benign	0.92	0.86	0.89
Malignant	0.84	0.91	0.87
Total	0.88	0.88	0.88

According to Table 2, the XGBoost model achieved an accuracy of 88.18%. While high precision (92%) was achieved for the benign class, high recall (91%) stood out for the malignant class. This suggests the model tends to reduce false positives in benign diagnoses and false negatives in malignant diagnoses. In a clinical context, a high recall value for the malignant class can be said to be advantageous as it reduces the risk of overlooking samples carrying the disease.

Table 3. Classification performance of Logistic Regression

Class	Precision	Recall	F1-Score
Benign	0.86	0.8	0.83
Malignant	0.78	0.84	0.81
Total	0.82	0.82	0.82

Table 3 indicate that, the Logistic Regression model provided an accuracy of 82.12%. While precision is high for the benign class, recall is relatively high for the malignant class. The overall performance is lower compared to XGBoost, suggesting that the assumption of a linear decision boundary is limited when dealing with complex color distributions.

Table 4. Classification performances of Random Forest

Class	Precision	Recall	F1-Score
Benign	0.92	0.81	0.86
Malignant	0.8	0.91	0.85
Total	0.86	0.86	0.86

According to Table 4, the Random Forest model achieved an accuracy of 85.61%. The model is strong in terms of precision (92%) in the benign class and has a high recall value (91%) in the malignant class. Although this tendency is similar to XGBoost, the overall accuracy remained slightly below that of XGBoost.

Table 5. Classification metrics of XGBoost, Logistic Regression and Random Forest

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	0.88	0.88	0.88	0.88
Logistic Regression	0.82	0.82	0.82	0.82
Random Forest	0.86	0.86	0.86	0.86

According to the comparative results in Table 5, the highest accuracy and most balanced performance were obtained by the XGBoost model (88%). Random Forest ranked second with 86% accuracy, while Logistic Regression showed the lowest performance with 82% accuracy. Particularly, the high recall values in the malignant class reveal that the XGBoost and Random Forest models are stronger compared to Logistic Regression.

Results and Conclusion

Studies on medical image processing reveal that classification-based approaches are widely used for the diagnosis of specific diseases. Moving from this point, the current study utilized an existing skin cancer dataset to classify images into benign and malignant categories using three different deep learning algorithms. Precision, Recall, and F1 Score were considered to determine the most successful of the three methods, and XGBoost was identified as the top-performing method with a precision of 88%. Compared with the other studies it can be said that its not a surprising result.

In addition to the network features used for deep learning, the image quality and volume of the data utilized for training and testing are also critical for model performance. Considering the results obtained from this study within its existing limitations, it is hoped that this research will serve as a guiding framework for future studies conducted by researchers in this field. Like many other studies performances of XGBoost is much better than the others, which causes the reason of recommending this model for similar studies.

References

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Chan, S., Reddy, V., Myers, B., Thibodeaux, Q., Brownstone, N., & Liao, W. (2020). Machine learning in dermatology: Current applications, opportunities, and limitations. *Dermatology and Therapy*, 10(2), 365–386.
<https://doi.org/10.1007/s13555-020-00372-0>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.
<https://doi.org/10.1145/2939672.2939785>
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Das, K., Cockerell, C. J., Patil, A., Pietkiewicz, P., Giulini, M., Grabbe, S., & Goldust, M. (2021). Machine learning and its application in skin cancer. *International Journal of Environmental Research and Public Health*, 18(24), 13409.
<https://doi.org/10.3390/ijerph182413409>
- Das, S., Dey A., Pal, A. & Roy N., (2015). Applications of Artificial Intelligence in Machine Learning: Review and Prospect. *International Journal of Computer Applications*. Volume 115, No. 9, pp. 31-41.
- Elston, D. M., Stratman, E. J., & Miller, S. J. (2016). Skin biopsy: Biopsy issues in specific diseases. *Journal of the American Academy of Dermatology*, 74(1), 1–16. <https://doi.org/10.1016/j.jaad.2015.07.017>
- Glazer, A. M., Winkelmann, R. R., Farberg, A. S., & Rigel, D. S. (2017). Analysis of trends in US melanoma incidence and mortality. *JAMA Dermatology*, 153(2), 225–226. <https://doi.org/10.1001/jamadermatol.2016.4512>
- Haddad, S., Weise, J. J., Wagenpfeil, S., Vogt, T., & Reichrath, J. (2025). Malignant melanoma: Vitamin D status as a risk and prognostic factor – Meta-analyses and systematic review. *Anticancer Research*, 45(1), 27–37.
<https://doi.org/10.21873/anticancer.16717>
- Han, J., Colditz, G. A., & Hunter, D. J. (2006). Risk factors for skin cancers: A nested case-control study within the Nurses' Health Study. *International Journal of Epidemiology*, 35(6), 1514–1521. <https://doi.org/10.1093/ije/dyl197>
- Hasan N, Imran M, Nadeem M, Jain D, Haider K, Moshahid Alam Rizvi M, Sheikh A., Kesharwani P., Jai G.K., & Ahmad F.J. (2023a). Formulation and development of novel lipid-based combinatorial advanced nanoformulation for effective treatment of non-melanoma skin cancer. *International Journal*

- of Pharmaceutics, 632, 122580.
<https://doi.org/10.1016/j.ijpharm.2022.122580>
- Hasan, N., Nadaf, A., Imran, M., Jiba, U., Sheikh, A., Almalki, W. H., Almuji, S. S., Mohammed, Y. H., Kesharwani, P., & Ahmad, F. J. (2023b). Skin cancer: Understanding the journey of transformation from conventional to advanced treatment approaches. *Molecular Cancer*, 22, 168.
<https://doi.org/10.1186/s12943-023-01854-3>
- Hernández-Orallo, J., Martínez-Plumed, F., Schmid, U., Siebers, M., & Dowe, D. L. (2016). Computer models solving intelligence test problems: Progress and implications. *Artificial Intelligence*, 230, 74–107.
<https://doi.org/10.1016/j.artint.2015.09.006>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118548387>
- Jiao, W., Hao, X., & Qin, C. (2021). The Image Classification Method with CNN-XGBoost Model Based on Adaptive Particle Swarm Optimization. *Information*, 12(4), 156. <https://doi.org/10.3390/info12040156>
- Koza, J. R., Bennett, F. H., Andre, D., & Keane, M. A. (1996). Automated design of both the topology and sizing of analog electrical circuits using genetic programming. In J. S. Gero, & F. Sudweeks (Eds.), *Artificial Intelligence in Design '96*. Springer. https://doi.org/10.1007/978-94-009-0279-4_9
- Kühl, N., Schemmer, M., Goutier, M. & Satzger G. (2022). Artificial intelligence and machine learning. *Electron Markets* 32, 2235–2244
<https://doi.org/10.1007/s12525-022-00598-0>
- Minko, T., Rodriguez-Rodriguez, L., & Pozharov, V. (2013). Nanotechnology approaches for personalized treatment of multidrug resistant cancers. *Advanced Drug Delivery Reviews*, 65,(13–14), 1880–1895.
<https://doi.org/10.1016/j.addr.2013.07.016>
- Moldovanu, S., Munteanu, D., & Sirbu, C. (2025). Impact on Classification Process Generated by Corrupted Features. *Big Data and Cognitive Computing*, 9(2), 45. <https://doi.org/10.3390/bdce9020045>
- Nielsen, D. (2016). Tree boosting with XGBoost – Why does XGBoost win “every” machine learning competition? (Master’s thesis). Norwegian University of Science and Technology. <https://doi.org/10.13140/RG.2.2.25736.57606>
- Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3–14. <https://doi.org/10.1080/00220670209598786>
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier

- for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Rogers, H. W., Weinstock, M. A., Feldman, S. R., & Coldiron, B. M. (2012). Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the U.S. population. *JAMA Dermatology*, 151(10), 1081–1086. <https://doi.org/10.1001/jamadermatol.2015.1187>
- Sathe, N. C., & Zito, P. M. (2025). Skin cancer. In StatPearls [Internet]. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK441949/>
- Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, *A Cancer Journal for Clinicians*, 73(1), 17–48. <https://doi.org/10.3322/caac.21763>
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>
- Tiwari, N., Kumar, D., Priyadarshani, A., Jain, G. K., Mittal, G., Kesharwani P. & Aggarwal G. (2023). Recent progress in polymeric biomaterials and their potential applications in skin regeneration and wound care management. *Journal of Drug Delivery Science and Technology*, 82, 104319. <https://doi.org/10.1016/j.jddst.2023.104319>
- Zhang, Z., Zhao, Y., Canes, A., Steinberg, D., & Lyashevskaya, O. (2020). Predictive analytics with gradient boosting in clinical medicine. *Annals of Translational Medicine*, 7(7), 152. <https://doi.org/10.21037/atm.2019.03.29> <https://www.kaggle.com/>