

# Kümeleme Analizleri ve Uygulamaları

Öğr.Gör. Ahu Cephe

Uygulamalı Tıpta Yapay Zeka Eğitimi

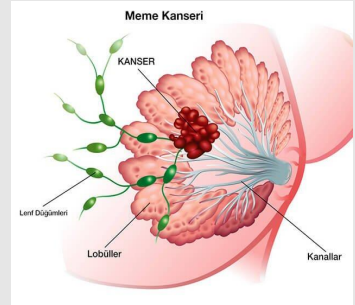
Erciyes Üniversitesi

15 Mayıs 2022

## Heterojen Bir Hastalığı Tanımak ve Tanımlamak

### Meme Kanseri

- Hastadan hastaya farklı türler
- Farklı biyolojik davranış
- Genç kadınlarda (40 yaşından küçük) daha hızlı seyretmekte iken, ileri yaş hastalarda (70 yaşından büyük) daha yavaş seyirli



## Heterojen Hastalıklar

- Kanser
- Felç
- Kardiyovasküler Hastalıklar
- Otizm
- Bulaşıcı Hastalıklar
- Otoimmün Hastalıklar
- Obezite
- Glokom
- ...

## Heterojen Bir Hastalığı Tedavi Etmek

- Heterojen hastalıklar tek bir hastalık olarak tanımlanamaz.
- Diagnostik, prognostik, terapötik moleküler biyobelirteçlerin tespiti doğru teşhis ve tedavi için son derece önemlidir.
- Hastalığın kimlik kartı hastalığın etki ettiği doku ve/veya organın moleküler yapısından gizlidir.

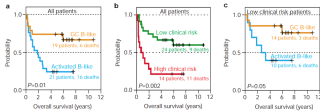
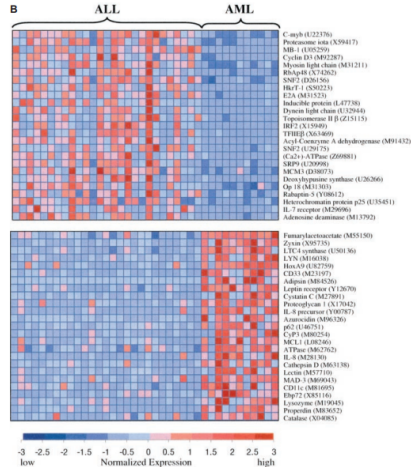
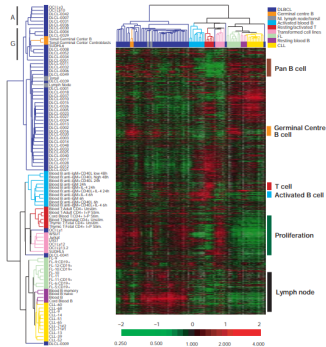
## Heterojen Hastalıklarla Karşılaşıldığında Ne Yapmalı?

Hastalıkları farklı alt türlere ayırmak

- Alt türlere göre farklı klinik denemeler oluşturmak
- Daha basit ve sade modeller oluşturmak
- Alt türlere özel klavuzlar oluşturmak
- Bireysel tıpta daha etkin ve verimli tedaviler uygulamak

# Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling

## Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring

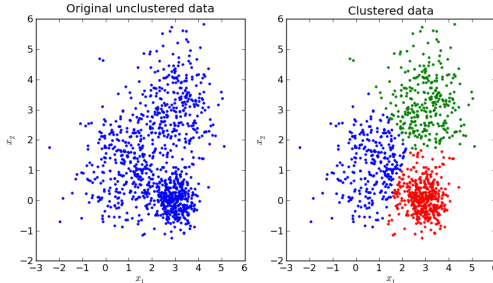


## Sınıflandırma ve Kümeleme Analizleri Arasındaki Farklar

	Clustering	Classification
<b>DEFINITION</b>	An unsupervised learning technique that group similar instances on the basis of features	A supervised learning technique that assign predefined tags to instances on the basis of features
<b>SUPERVISION</b>	Unsupervised learning technique	Supervised learning technique
<b>TRAINING DATASET</b>	Does not use training dataset	Uses a training dataset
<b>LABELS</b>	There are no labels	There are labels for training data
<b>PROCESS</b>	Uses statistical concepts, dataset is split into subsets with similar features	Uses the algorithms to categorize the new data according to the observations of the training set
<b>AIM</b>	Group a set of objects in order to find whether there is any relationship between them	Find to which class a new object belongs to from the set of predefined classes

## Kümeleme Analizi

- Amaç: Benzer gözlemleri aynı gruplara (kümelere), farklılık gösteren gözlemleri farklı gruplara atamak
- Kümeleme analizleri, kendi içinde benzer (homojen), kendi aralarında farklı (heterojen) gruplardan oluşan bir yapı oluşturur.





## Sağlık Alanında Kümeleme Analizleri

- Hastalık alt türlerini keşfetmek
- Hastalık evrelerini bulmak

**Cross-sectional Subtypes**



Figure 1. Subtypes of a disease by phenotype using cross-sectional biodata; for example, a single measure of blood sugar or a medical image. An example would be diabetes which is subtyped into "Type I" and "Type II".  
Image created by author.

**Stages**

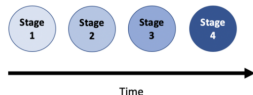


Figure 2. Grouping disease by stage, using longitudinal data which measures progression over time. An example would be cancer which is usually grouped into stages 1-4. Image created by Author.

## Kümeleme Analizinin Amaçları:

- Veriyi anlamak
- Veri görselleştirmek
- Boyut indirgemek

## Kümeleme Analizi Adımları:

- Veri matrisi oluşturma ve veri ön işleme
- Benzerlik veya uzaklık matrislerinin oluşturulması
- Küme sayısının belirlenmesi
- Kümeleme yönteminin uygulanması
- Küme geçerliliği ve güvenilirliği