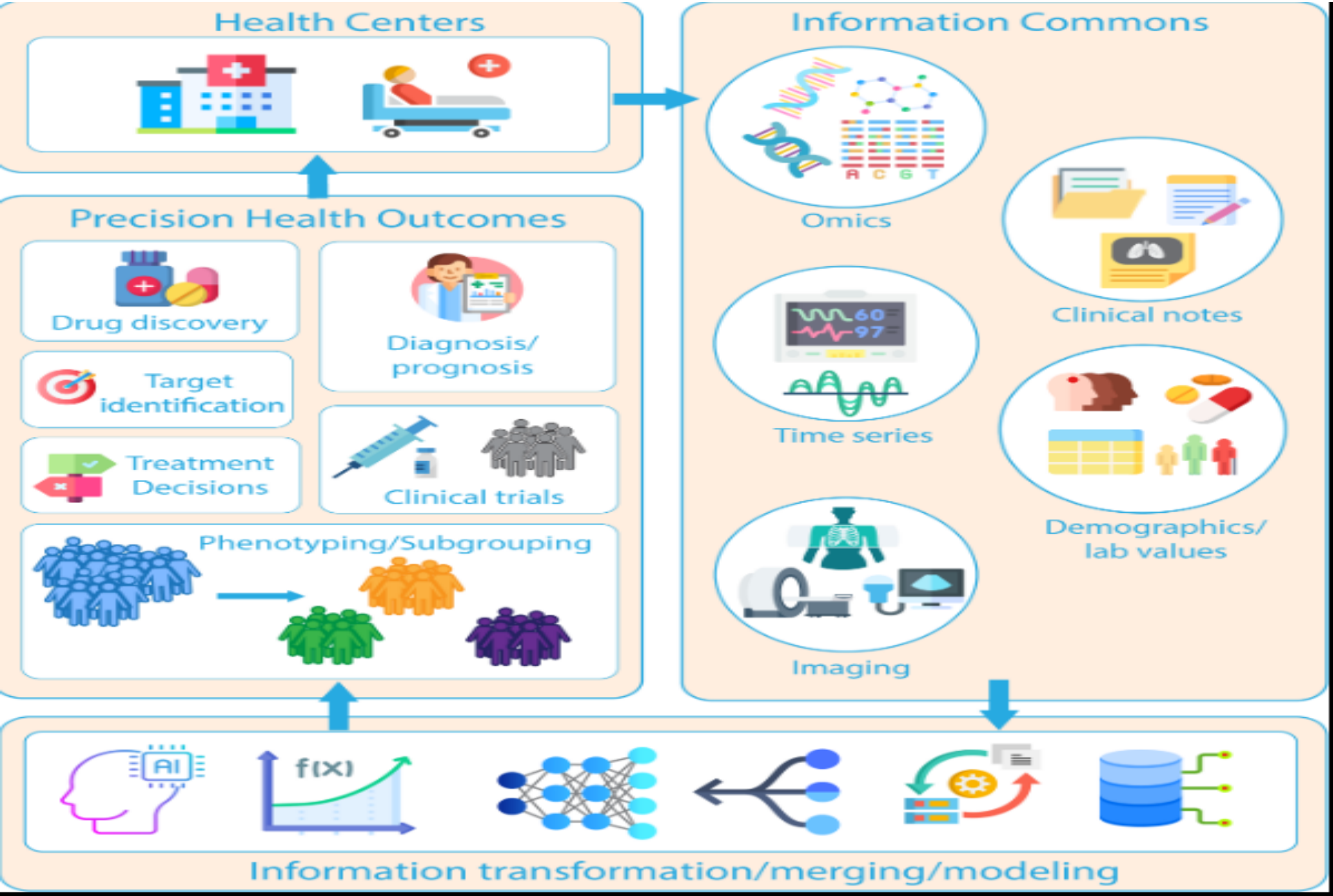
Clinical Data Mining

Lecture Three

By

Assist. Lect. Zainab M. Alameen

2025 - 2026


Health Centers
Information Commons
Omics
Clinical notes
Time series
Demographics/
lab values
Imaging
Precision Health Outcomes
Drug discovery
Diagnosis/
prognosis
Target
identification
Treatment
Decisions
Clinical trials
Phenotyping/Subgrouping
AI
f(x)
Information transformation/merging/modeling

Data Preprocessing and Cleaning I

Lecture Keys:

- Introduction.
- What is Data Preprocessing.
- Data Before preprocessing.
- Why Preprocessing is Important in Clinical Data Mining.
- Main Steps of Data Preprocessing.
- Challenges in Preprocessing Clinical Data.

Introduction:

- In real-world clinical settings, data are often incomplete, inconsistent, and noisy. Laboratory results may have missing values, patient records may contain errors, and sensor data can include random fluctuations.

What is Data Preprocessing:

- **Data preprocessing** is the essential first step to improve data quality and ensure reliable mining outcomes.
- Without proper preprocessing, any analytical or predictive model may give misleading results.

Data Before preprocessing:

Data can contain:

- Missing values (e.g., unrecorded lab test results)
- Inconsistent formats (e.g., temperature in °C vs °F)
- Noisy measurements (e.g., blood pressure sensor errors)

Why Preprocessing is Important in Clinical Data Mining

Preprocessing ensures:

- Improved data quality and accuracy
- Better model performance (e.g., classification, clustering, regression)
- Meaningful and reproducible results
- Regulatory and clinical reliability of medical decisions.

Main Steps of Data Preprocessing

1. Data Cleaning

- Handling missing values (deletion, imputation, interpolation)
- Identifying and correcting errors (e.g., negative age, unrealistic blood pressure)
- Removing duplicate or redundant records
- Dealing with noisy data using smoothing techniques.

Main Steps of Data Preprocessing

2. Data Integration

- Combining data from multiple clinical sources:
  - Hospital information systems (HIS)
  - Laboratory Information Systems (LIS)
  - Radiology Information Systems (RIS)
  - Wearable sensors and patient monitoring devices
- Handling schema integration and data format inconsistencies.

Main Steps of Data Preprocessing:

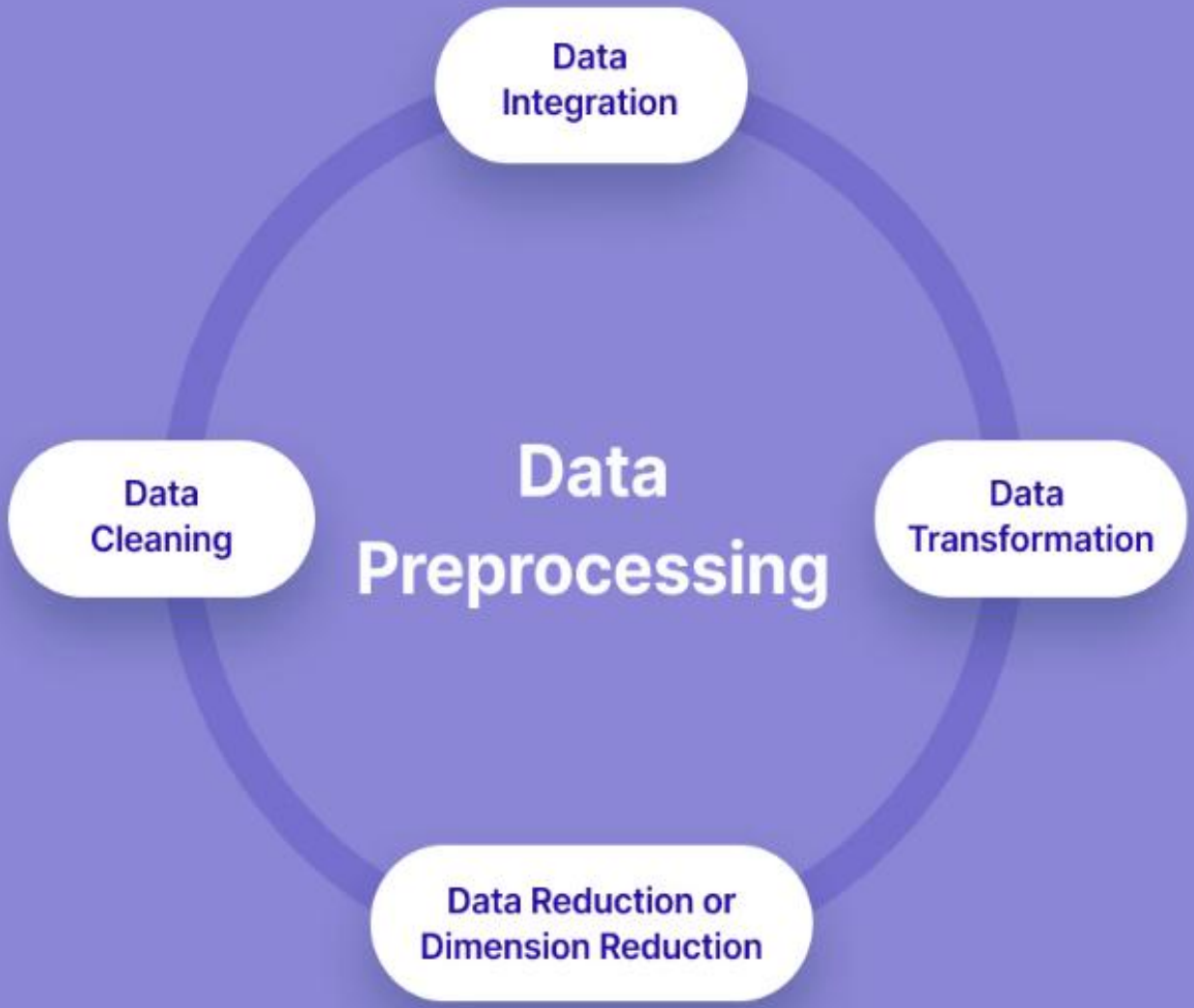
3. Data Transformation

- Normalization (e.g., bringing lab values into common units)
- Aggregation (e.g., daily average heart rate)
- Encoding categorical data (e.g., gender, diagnosis codes).

Main Steps of Data Preprocessing:

4. Data Reduction:

- Simplifying the dataset without losing key information (feature selection, dimensionality reduction).


Data Integration
Data Transformation
Data Reduction or Dimension Reduction
Data Cleaning
Data Preprocessing

Challenges in Preprocessing Clinical Data

- Patient privacy and confidentiality.
- High dimensionality of data (genomic, lab tests, imaging).
- Real-time data streams (e.g., ICU monitoring).
- Interoperability between systems.

The End

Thanks for your listening