Smart Hospital – Module 2

22/04/2024

Precision Medicine

Alessandra Marasi: alessandra.marasi@polimi.it

Francesca Fati: francesca.fati@polimi.it

Sofia Breschi: sofia.breschi@polimi.it

Eriberto Andrea Franchi: eribertoandrea.franchi@polimi.it

POLITECNICO
MILANO 1863

Agenda

- Precision Medicine
- Data in Healthcare
- Medical AI: Model-centric vs Data-centric AI
- Data Processing
- Hands-On: Heart Disease Dataset

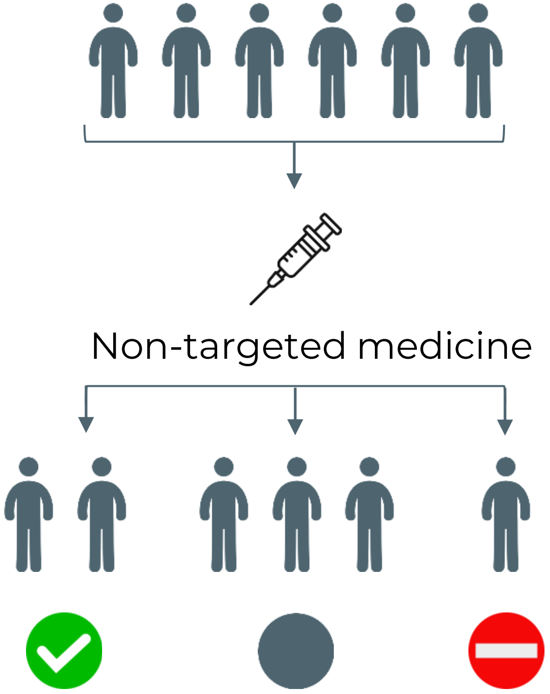
Agenda

- **Precision Medicine**
- Data in Healthcare
- Medical AI: Model-centric vs Data-centric AI
- Data Processing
- Hands-On: Heart Disease Dataset

POLITECNICO
MILANO 1863

Precision Medicine: Definition

Traditional approach



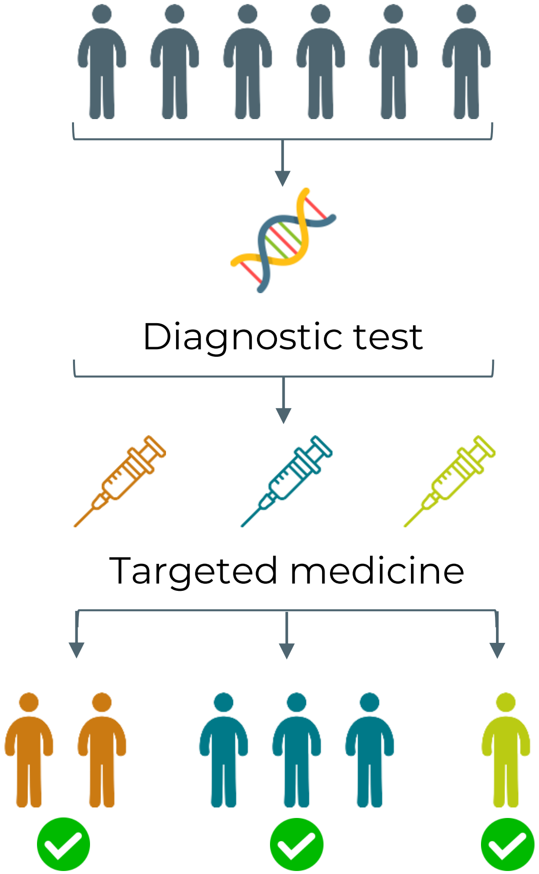
Non-targeted medicine

Positive effect

No effect

Negative effect

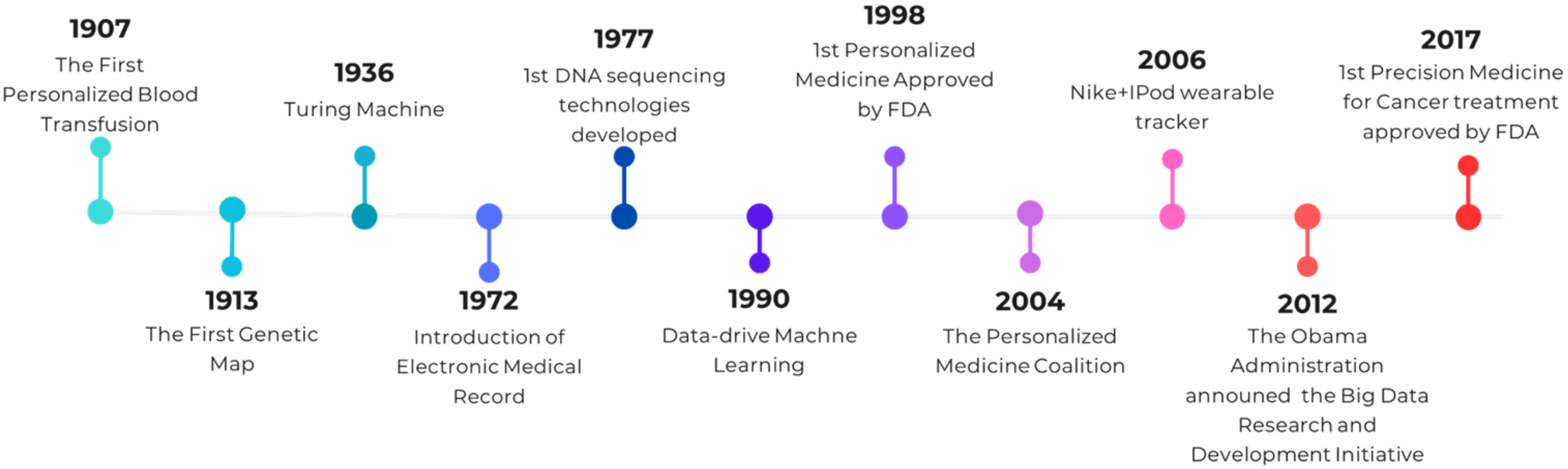
Personalised approach



Diagnostic test

Targeted medicine

Precision Medicine: Timeline



1907
The First Personalized Blood Transfusion

1913
The First Genetic Map

1936
Turing Machine

1972
Introduction of Electronic Medical Record

1977
1st DNA sequencing technologies developed

1990
Data-drive Machne Learning

1998
1st Personalized Medicine Approved by FDA

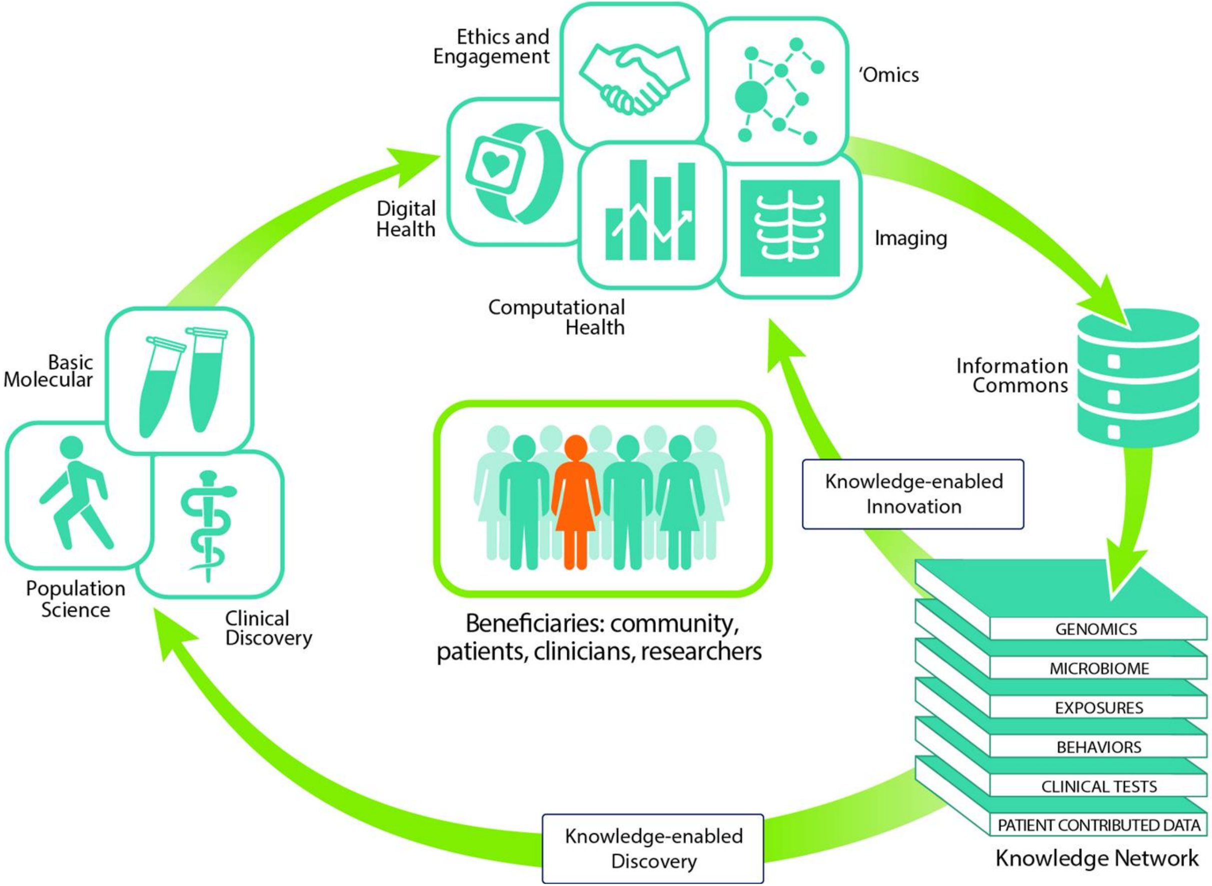
2004
The Personalized Medicine Coalition

2006
Nike+IPod wearable tracker

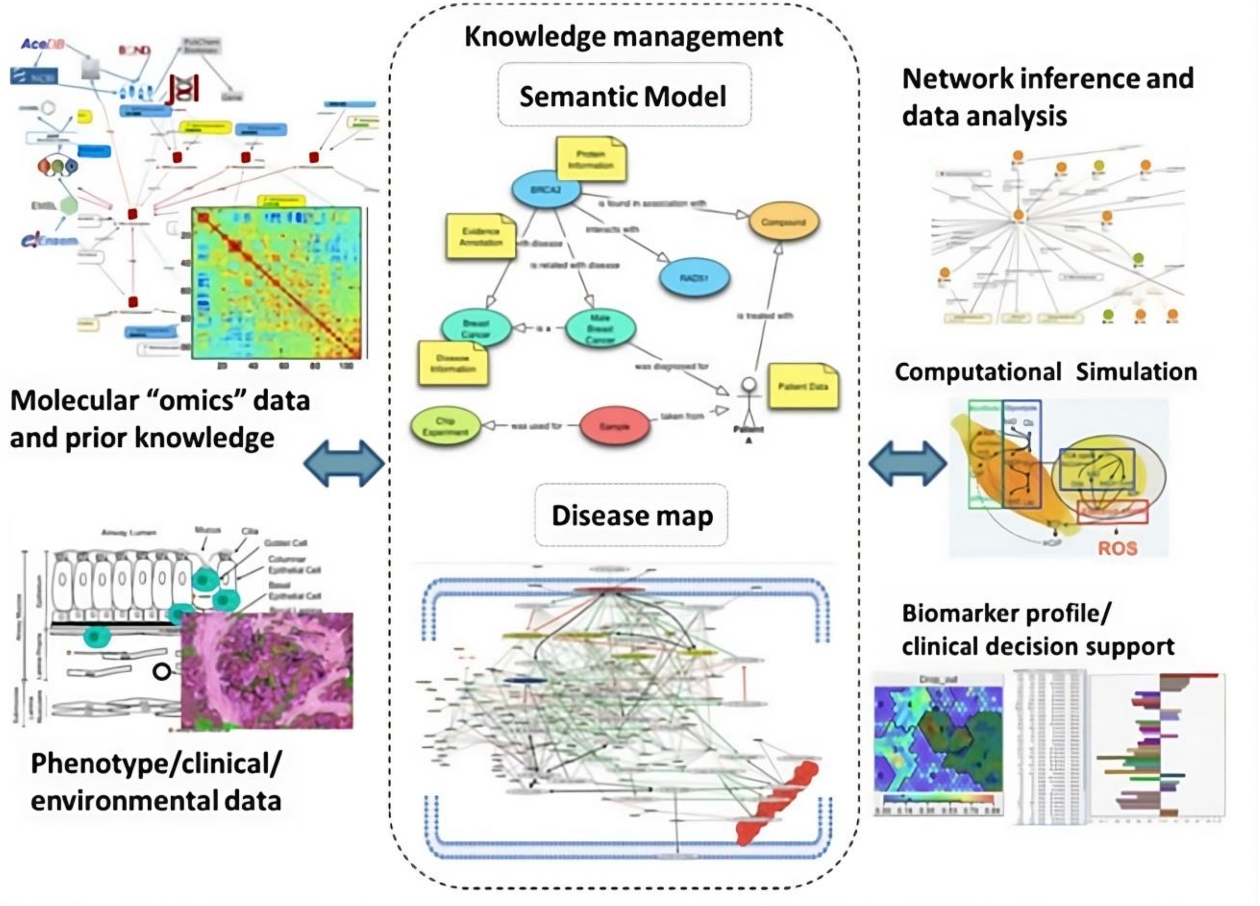
2012
The Obama Administration announed the Big Data Research and Development Initiative

2017
1st Precision Medicine for Cancer treatment approved by FDA

Precision Medicine: Ecosystem



Precision Medicine: Ecosystem



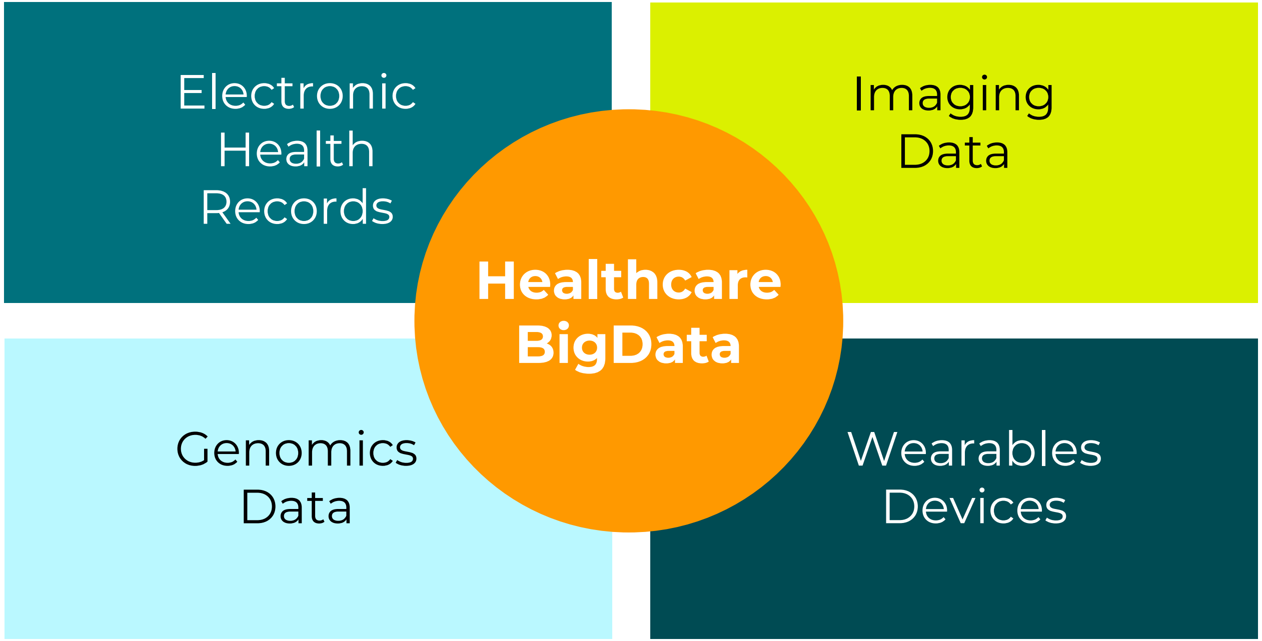
Agenda

- Precision Medicine
- **Data in Healthcare**
- Medical AI: Model-centric vs Data-centric AI
- Data Processing
- Hands-On: Heart Disease Dataset

Data in Healthcare

It’s all about data.

Large data sets (Big Data) to find insights, trends, and patterns.



Electronic Health Records

Imaging Data

Healthcare BigData

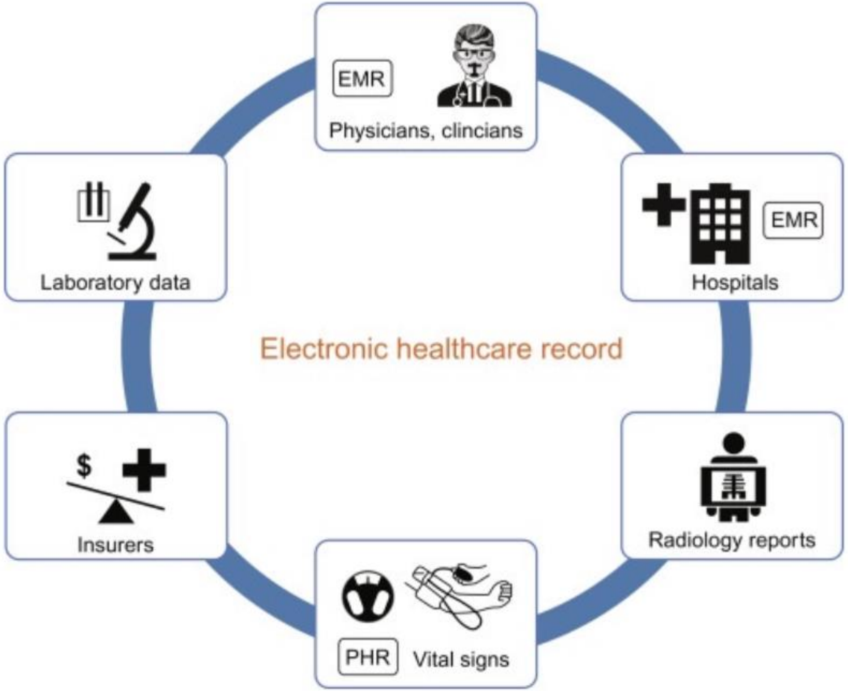
Genomics Data

Wearables Devices

Data in Healthcare

Electronic Health Records

- Patient demographics
- Medical history
- Medication and allergies
- Laboratory test results
- Radiology images
- Vital signs
- Personal statistics (e.g., age and weight)

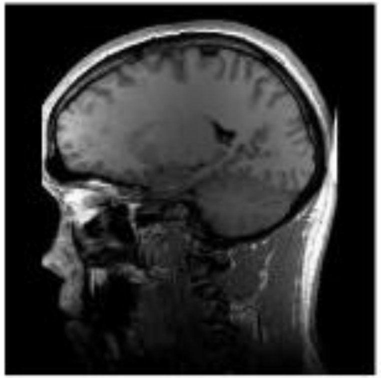


Data in Healthcare

Imaging Data

- X-rays
- Magnetic Resonance Imaging (MRI)
- Computed Tomography (CT)
- Ultrasound
- Endoscopy

CT



MRI

X-rays

Ultrasound

Data in Healthcare

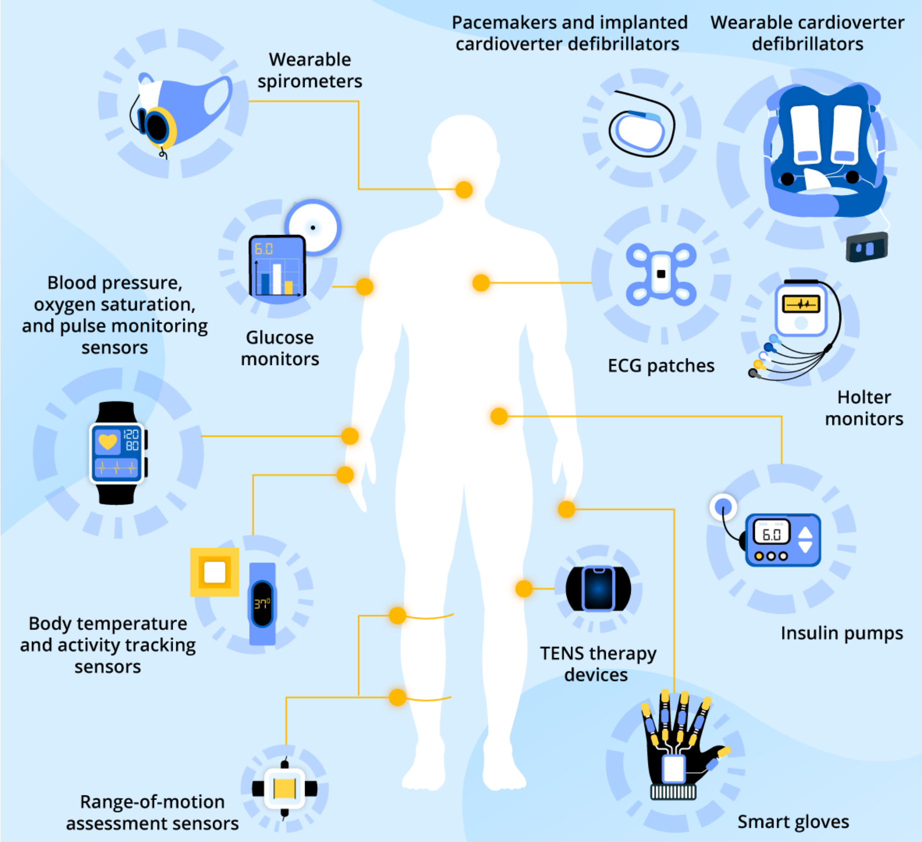
Genomics Data

- Whole Genome Sequencing (WGS) Data
- Whole Exome Sequencing (WES) Data
- Transcriptome Sequencing (RNA-Seq) Data
- Methylation Sequencing Data
- Single-Cell Sequencing Data

Data in Healthcare

Wearables Devices

- Activity Trackers
- Smart Health Watches
- Wearable ECG Monitors
- Blood Pressure Monitors
- Continuous Glucose Monitors
- Wearable Biosensors



Wearable spirometers
Pacemakers and implanted cardioverter defibrillators
Wearable cardioverter defibrillators
Blood pressure, oxygen saturation, and pulse monitoring sensors
Glucose monitors
ECG patches
Holter monitors
Body temperature and activity tracking sensors
TENS therapy devices
Insulin pumps
Range-of-motion assessment sensors
Smart gloves

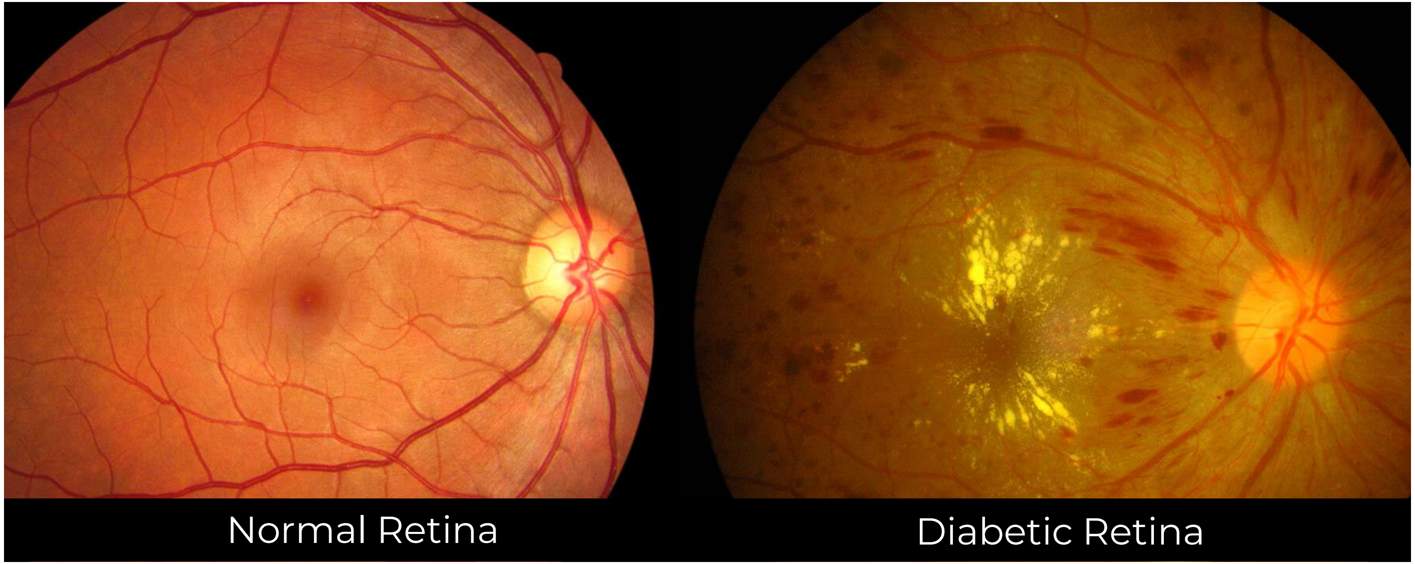
Agenda

- Precision Medicine
- Data in Healthcare
- **Medical AI - Model-centric vs Data-centric AI**
- Data Processing
- Hands-On: Heart Disease Dataset

Medical AI : Diabetic Retinopathy

Diabetic retinopathy

- Over 420 million people with diabetes globally
- High blood sugar levels associated with diabetes can lead to damage to the blood vessels of the retina.
- Early stages may cause mild vision problems. It can lead to blindness.



Medical AI : Diabetic Retinopathy

Diagnosis

- Ophthalmoscopy or Fundus Photography

Treatment

- Laser Surgery
- Vitrectomy

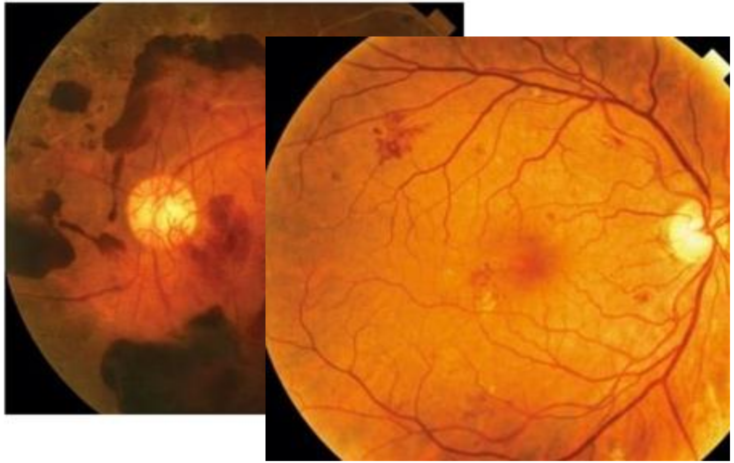
Challenges:

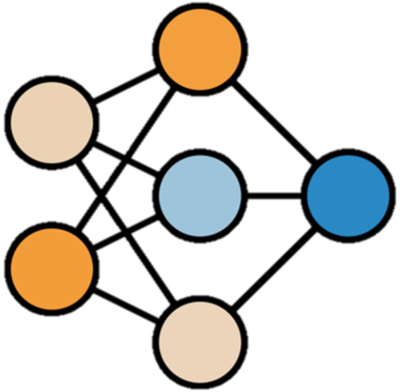
- Early Detection: no noticeable symptoms, delaying diagnosis and treatment
- Variability: signs might be less apparent in some individuals
- Interpretation of Diagnostic Tests: the interpretation of fluorescein angiography requires expertise and experience, leading to under or over-diagnosis
- Differential Diagnosis: other conditions can mimic the signs of diabetic retinopathy, such as retinal vein occlusions or age-related macular degeneration, complicating the diagnosis

Medical AI: ARDA

Automated Retinal Disease Assessment (ARDA)

- AI-based system which interprets retinal scans to detect diabetic retinopathy
- Large team of ophthalmologists
- Manually reviewing **more than 100,000** de-identified retinal scans





Diabetic Retinopathy:
Yes/No

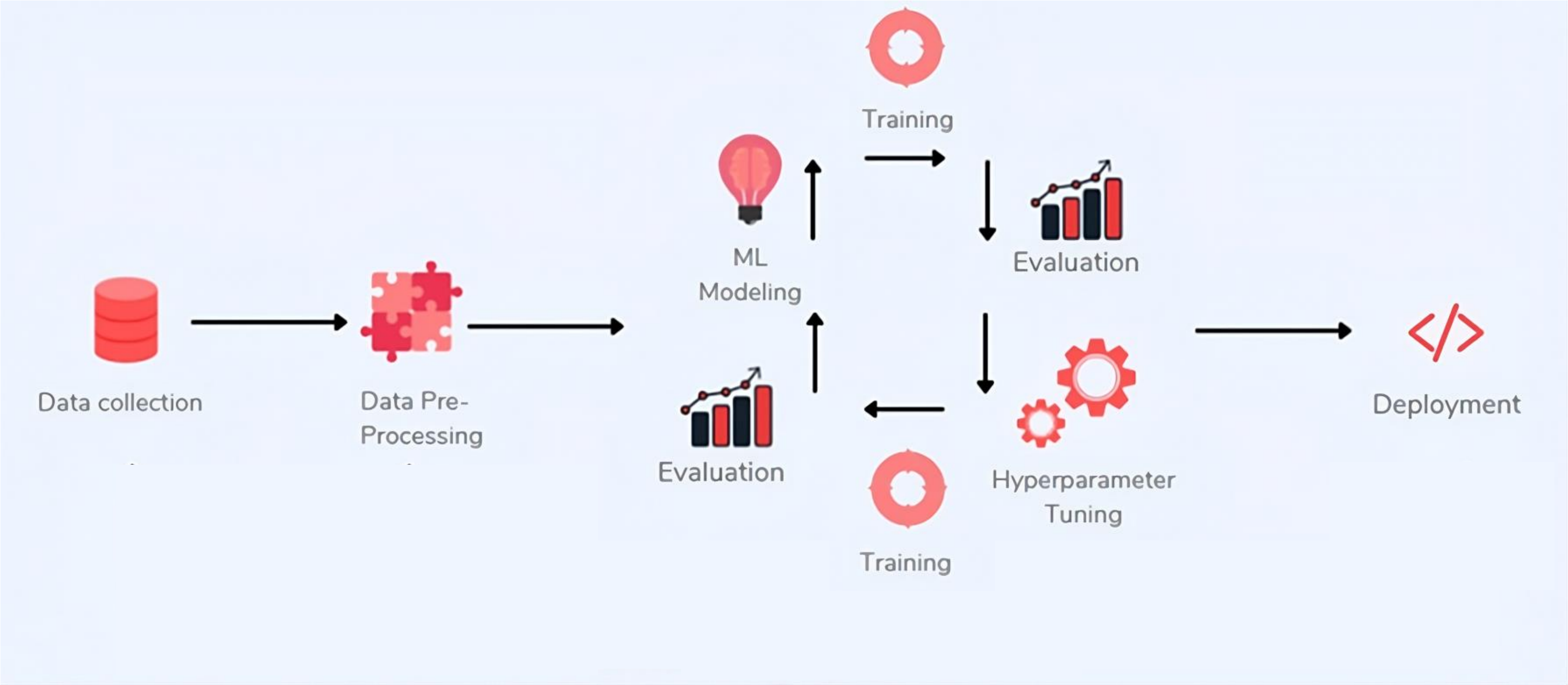
Input: Retinal Scans

AI model

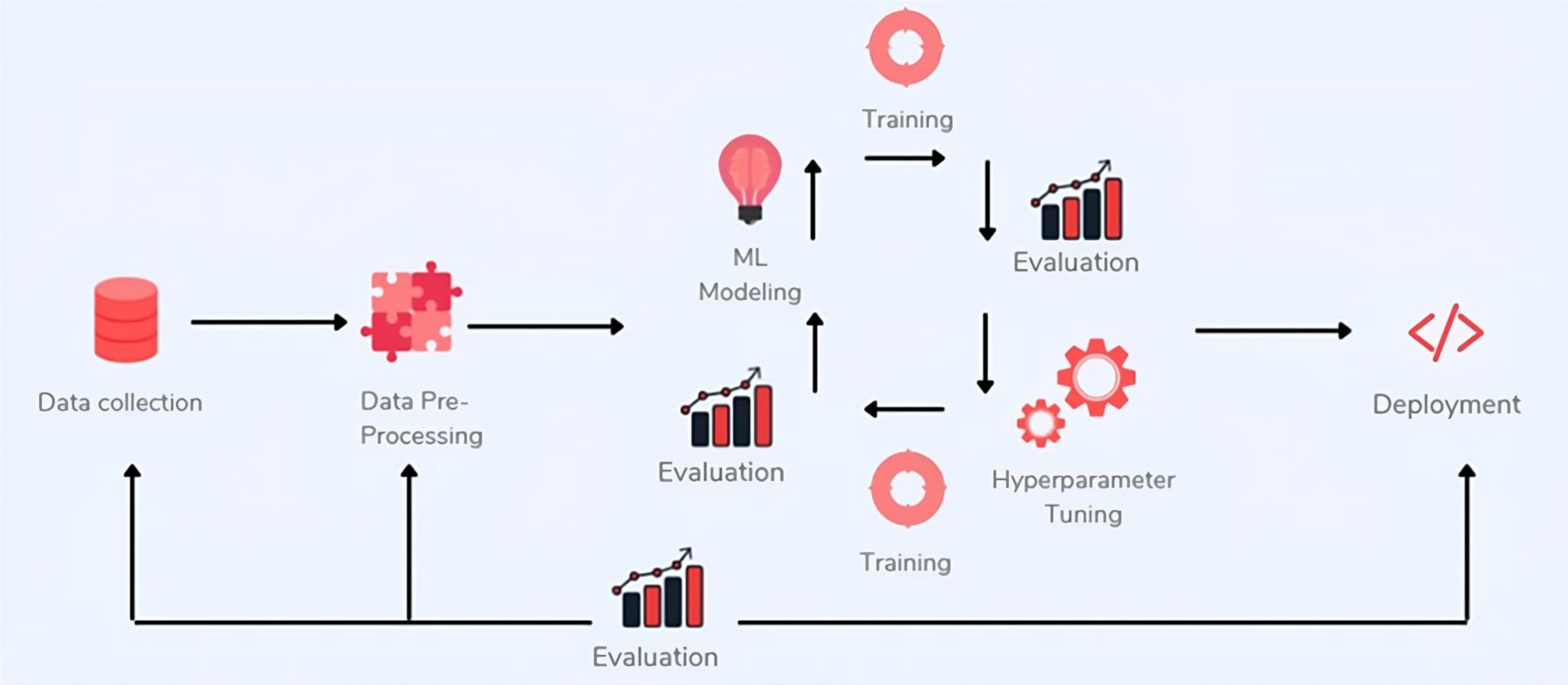
Output: DR Detection

https://health.google/caregivers/arda/

Model-centric AI



Data-centric AI



Agenda

- Precision Medicine
- Data in Healthcare
- Medical AI: Model-centric vs Data-centric AI
- **Data Processing**
- Hands-On: Heart Disease Dataset

Data Processing

i. Data Visualization

ii. Handle Missing Data

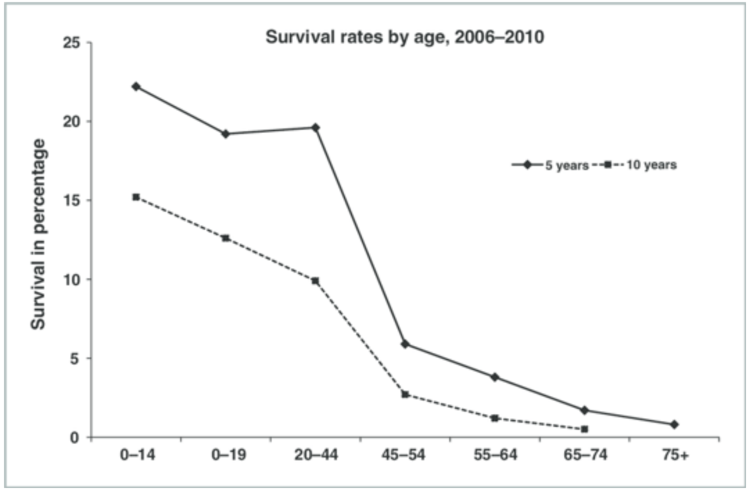
iii. Data Analytics

iv. Data Augmentation

Data Processing

Clinical Case: Glioblastomas

- The **most common primary malignant** brain tumor
- Annual incidence ranging from 6 to 10 cases per 100,000 population
- Glioblastomas (Grade IV) are the most aggressive primary brain tumor
- **Median survival around 12-15 months**



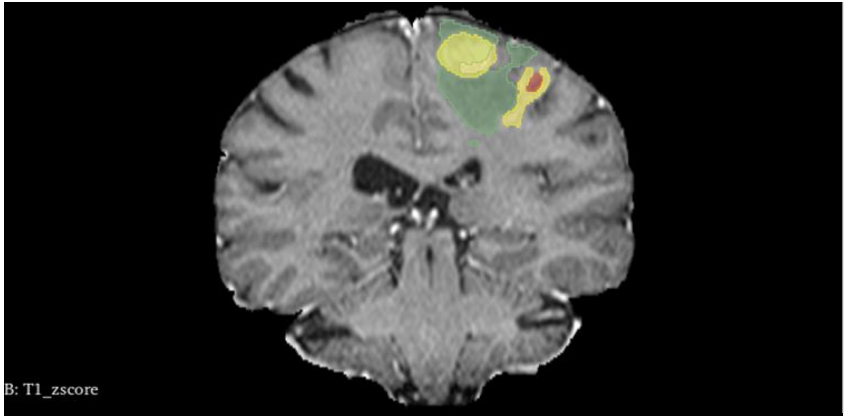
Data Processing

Clinical Case: Glioblastomas

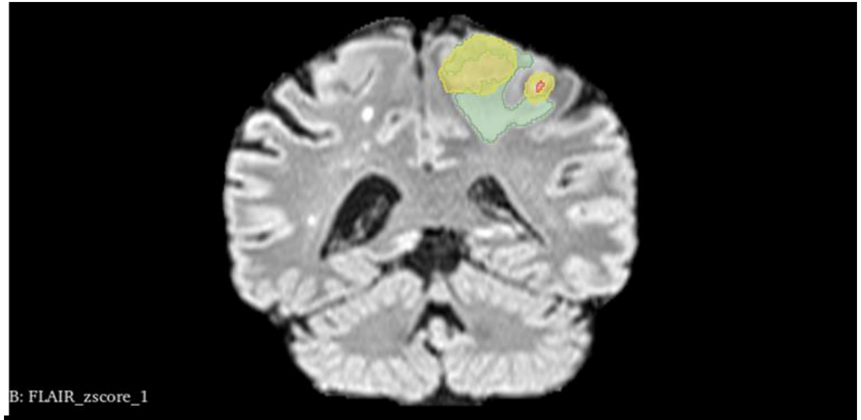
- Diagnosis and evaluation of treatment response: **Medical Resonance Imaging (MRI)**
  - T1 MRI
  - T1 contrast-enhanced MRI
  - T2 MRI
  - FLAIR MRI
- **Surgery** is the first-line treatment followed by radiotherapy
- Radiotherapy and Chemotherapy are also used when the surgical removal or the total resection is not possible

Data Processing

Clinical Case: Glioblastomas



T1 contrast-enhanced MRI



FLAIR MRI

Tumor Enhancement

Tumor Necrosis

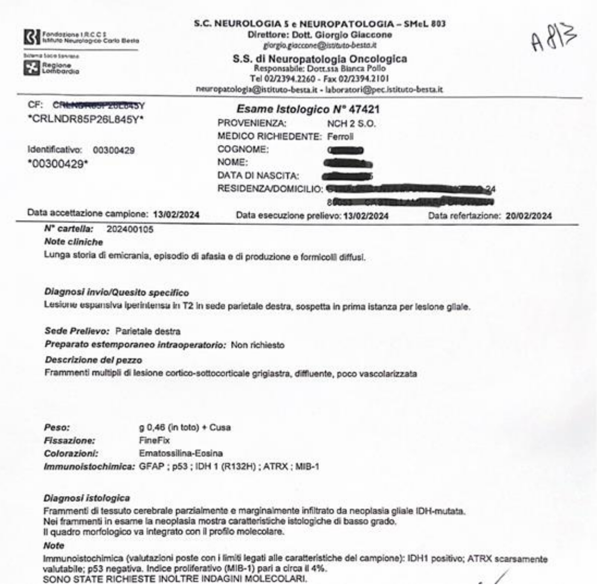
FLAIR Hyperintensities

Data Processing

Data Visualization | Handle Missing Data | Data Analytics | Data Augmentation

Clinical Case: Glioblastomas

- Clinical Database



Fondazione I.R.C.C.S. Istituto Neurologico Carlo Besta

Sistema Socio Sanitario Regione Lombardia

S.C. NEUROLOGIA 5 e NEUROPATOLOGIA – SMeL 803
Direttore: Dott. Giorgio Giaccone
giorgio.giaccone@istituto-besta.it
S.S. di Neuropatologia Oncologica
Responsabile: Dott.ssa Bianca Pollo
Tel 02/2394.2260 - Fax 02/2394.2101
neuropatologia@istituto-besta.it - laboratori@pec.istituto-besta.it

A 813

CF: [redacted]
CRLNDR85P26L845Y

Identificativo: 00300429
00300429

Esame Istologico N° 47421

PROVENIENZA: NCH 2 S.O.
MEDICO RICHIEDENTE: Ferroli
COGNOME: [redacted]
NOME: [redacted]
DATA DI NASCITA: [redacted]
RESIDENZA/DOMICILIO: [redacted]

Data accettazione campione: **13/02/2024** Data esecuzione prelievo: **13/02/2024** Data refertazione: **20/02/2024**

N° cartella: 202400105

Note cliniche

Lunga storia di emicrania, episodio di afasia e di produzione e formicolii diffusi.

Diagnosi invio/Quesito specifico

Lesione espansiva iperintensa in T2 in sede parietale destra, sospetta in prima istanza per lesione gliale.

Sede Prelievo: Parietale destra

Preparato estemporaneo intraoperatorio: Non richiesto

Descrizione del pezzo

Frammenti multipli di lesione cortico-sottocorticale grigiastra, diffluente, poco vascolarizzata

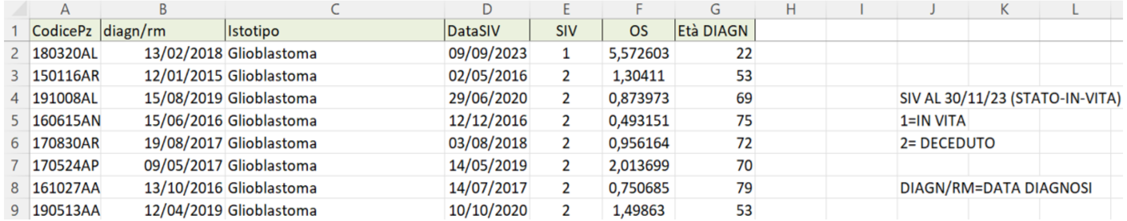
Peso: g 0,46 (in toto) + Cusa
Fissazione: FineFix
Colorazioni: Ematossilina-Eosina
Immunoistochimica: GFAP ; p53 ; IDH 1 (R132H) ; ATRX ; MIB-1

Diagnosi istologica

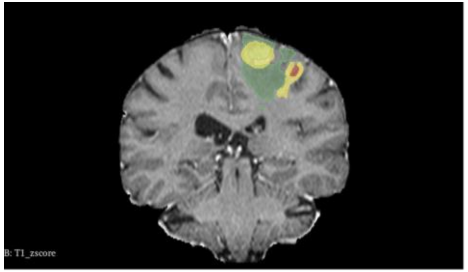
Frammenti di tessuto cerebrale parzialmente e marginalmente infiltrato da neoplasia gliale IDH-mutata.
Nei frammenti in esame la neoplasia mostra caratteristiche istologiche di basso grado.
Il quadro morfologico va integrato con il profilo molecolare.

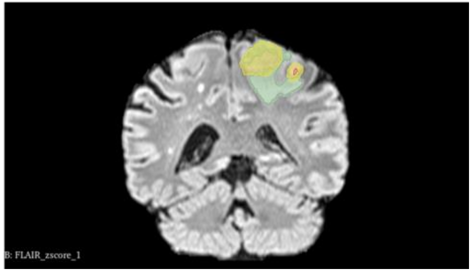
Note

Immunoistochimica (valutazioni poste con i limiti legati alle caratteristiche del campione): IDH1 positivo; ATRX scarsamente valutabile; p53 negativa. Indice proliferativo (MIB-1) pari a circa il 4%.
SONO STATE RICHIESTE INOLTRE INDAGINI MOLECOLARI.



	A	B	C	D	E	F	G	H	I	J	K	L
1	CodicePz	diagn/rm	Istotipo	DataSIV	SIV	OS	Età DIAGN					
2	180320AL	13/02/2018	Glioblastoma	09/09/2023	1	5,572603	22					
3	150116AR	12/01/2015	Glioblastoma	02/05/2016	2	1,30411	53					
4	191008AL	15/08/2019	Glioblastoma	29/06/2020	2	0,873973	69			SIV AL 30/11/23 (STATO-IN-VITA)		
5	160615AN	15/06/2016	Glioblastoma	12/12/2016	2	0,493151	75			1=IN VITA		
6	170830AR	19/08/2017	Glioblastoma	03/08/2018	2	0,956164	72			2= DECEDUTO		
7	170524AP	09/05/2017	Glioblastoma	14/05/2019	2	2,013699	70					
8	161027AA	13/10/2016	Glioblastoma	14/07/2017	2	0,750685	79			DIAGN/RM=DATA DIAGNOSI		
9	190513AA	12/04/2019	Glioblastoma	10/10/2020	2	1,49863	53					





Agenda

- Precision Medicine
- Data in Healthcare
- Medical AI: Model-centric vs Data-centric AI
- Data Processing
- **Hands-On: Heart Disease Dataset**

Hands-On: Heart Disease Dataset

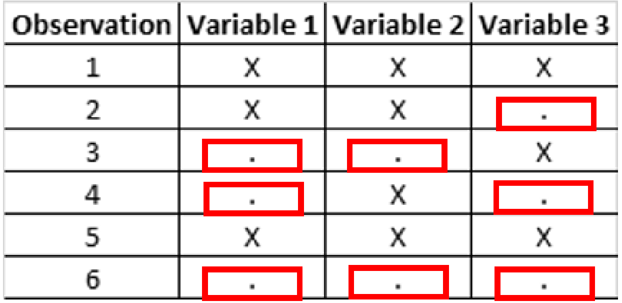
Data Visualization

https://tinyurl.com/3ywaw4wr

Data Processing

Missing Data: observations that we planned to register but we could not

- □ Some patients lost the follow-up visits
- □ Partially-filled medical questionnaires
- □ Incomplete medical records
- □ Loss of data during transfer
- □ Limited availability for specific tests or assessments



Observation	Variable 1	Variable 2	Variable 3
1	X	X	X
2	X	X	.
3	.	.	X
4	.	X	.
5	X	X	X
6	.	.	.

🟥 MISSING DATA

Data Processing

Data Visualization **Handle Missing Data** Data Analytics Data Augmentation

Missing Data: observations that we planned to register but we could not

- x Reduced reliability of study findings
- x Increase risk of bias
- x Reduced statistical power
- x Some populations may be under-represented

Data Processing

Missing Mechanisms

Missing Completely at Random (MCAR)

There is no statistically significant difference between incomplete and complete cases □ no relationship between the missing data to any other measured variables and itself

Example: In a clinical trial with 150 participants testing a new antidepressant, 7 participants have missing post-treatment depression scores. These missing values are randomly distributed across various demographic groups (age, gender, etc.)

Missing Mechanisms

Missing Random (MAR)

Dependency of a data point to be missing on some measured variables of the dataset and not to itself.

Example: *In a dataset of 500 participants, missing cognitive test scores are observed more frequently among individuals aged 65 and above. Missingness based on age, a different variable.*

Missing Mechanisms

Missing Not a Random (MNAR)

The missingness of a variable depends on the variable itself.

Example: In a clinical trial assessing the effectiveness of a weight loss program, participants who are not losing weight may avoid follow-up appointments. Out of 200 participants enrolled, 30 participants with missing weight data consistently report higher levels of dissatisfaction with the program.

Data Processing

Strategies for Handling Missing Data

- **Deletion** Methods
- **Imputation** Methods

1. Identify the reason
2. Understand the data distribution
3. Choose for the best strategy

Data Processing

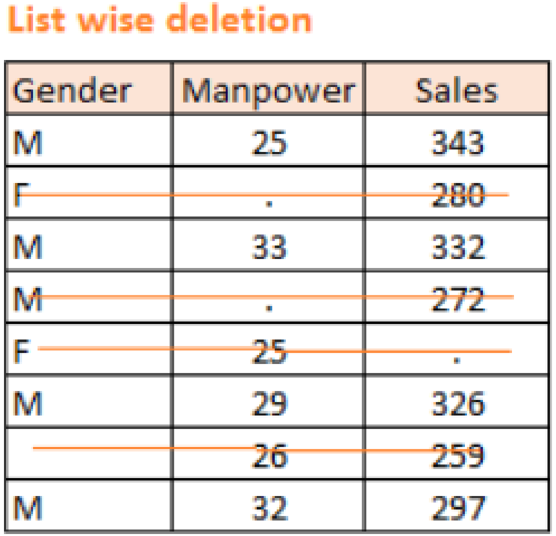
Strategies for Handling Missing Data

Deletion Methods: **Listwise**

Only analyse cases with available data on each variable

✓ Simplicity and comparability across all analysis

✗ Reduce of statistical power (decreased sample size), do not use all information



List wise deletion

Gender	Manpower	Sales
M	25	343
~~F~~	~~.~~	~~280~~
M	33	332
~~M~~	~~.~~	~~272~~
~~F~~	~~25~~	~~.~~
M	29	326
	~~26~~	~~259~~
M	32	297

Data Processing

Strategies for Handling Missing Data

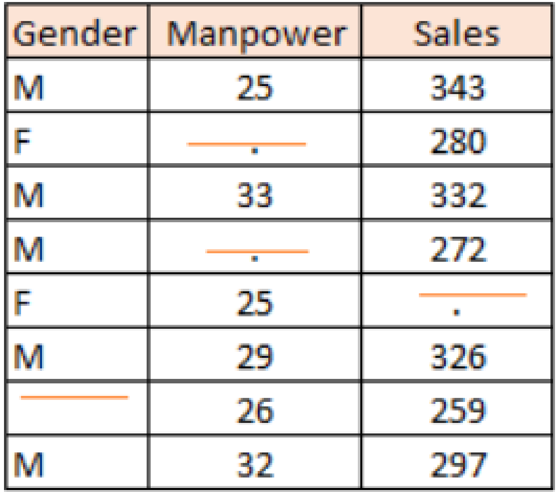
Deletion Methods: **Pairwise**

Analysis of all cases where the variable of interest is present

✓ Keeping of as many cases as possible, use of all information

✗ Cannot compare analyses due to different sample size at each time, sample size varies for each variable

Pair wise deletion



Gender	Manpower	Sales
M	25	343
F	~~.~~	280
M	33	332
M	~~.~~	272
F	25	~~.~~
M	29	326
	26	259
M	32	297

Data Processing

Strategies for Handling Missing Data

Imputations Methods: **Random Sample from a "Reasonable" Distribution**

i. Data distribution (Normal, Bernoulli,..) identification and its parameters

ii. Replace missing values with random draws from the distribution

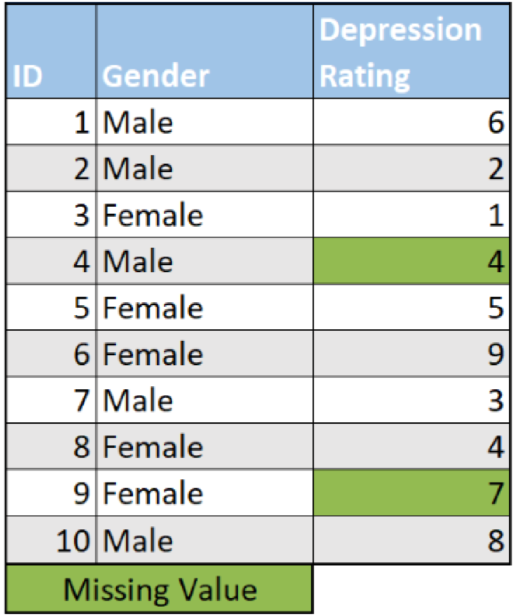
✓ Data distribution is preserved

✗ The identification of data distribution may not be reliable

Data Processing

Strategies for Handling Missing Data

- **Imputations** Methods: **Random Sample from a "Reasonable" Distribution**



ID	Gender	Depression Rating
1	Male	6
2	Male	2
3	Female	1
4	Male	🟩 4
5	Female	5
6	Female	9
7	Male	3
8	Female	4
9	Female	🟩 7
10	Male	8

🟩 Missing Value

Data Processing

Strategies for Handling Missing Data

Imputations Methods: **Mean/Mode Substitution**

i. Compute mean/median/mode values from the dataset

ii. Replace missing values the sample mean/median/mode

✓ Keeping of as many cases as possible, use of all information

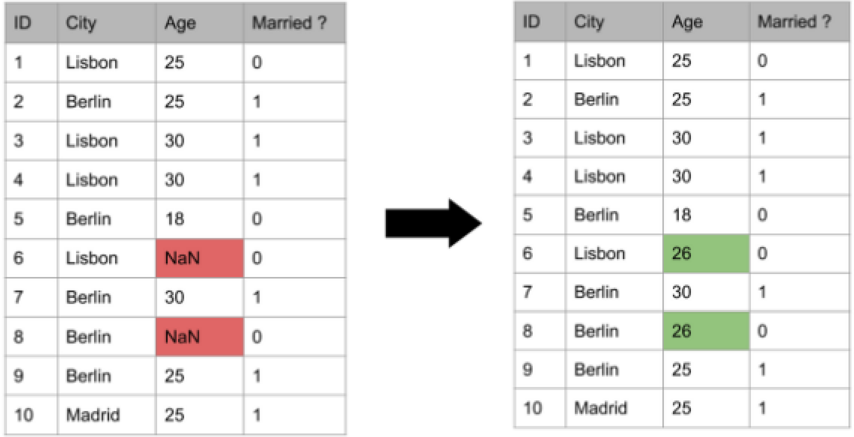
✗ Variability reduction

Data Processing

Strategies for Handling Missing Data

- **Imputations** Methods: **Mean/Mode Substitution**

Average_Age = 26.0



ID	City	Age	Married ?
1	Lisbon	25	0
2	Berlin	25	1
3	Lisbon	30	1
4	Lisbon	30	1
5	Berlin	18	0
6	Lisbon	NaN	0
7	Berlin	30	1
8	Berlin	NaN	0
9	Berlin	25	1
10	Madrid	25	1

ID	City	Age	Married ?
1	Lisbon	25	0
2	Berlin	25	1
3	Lisbon	30	1
4	Lisbon	30	1
5	Berlin	18	0
6	Lisbon	26	0
7	Berlin	30	1
8	Berlin	26	0
9	Berlin	25	1
10	Madrid	25	1

Data Processing

Strategies for Handling Missing Data

Imputations Methods: **Deterministic Regression Imputation**

Replace missing values with predicted scores from regression

i. Select a regression model (linear, polynomial, logistic, etc.)

$$y_{2,i} = \beta_0 + \beta_1 y_{1,i}$$

ii. Estimate regression parameters (β_0, β_1) from data

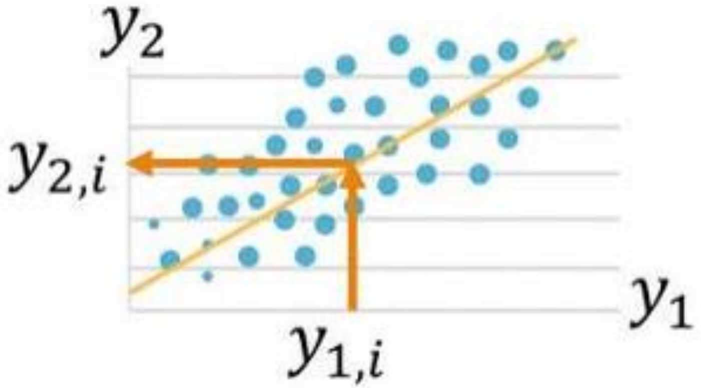
iii. Replace missing data using regression estimators

Data Processing

Strategies for Handling Missing Data

- **Imputations** Methods: **Deterministic Regression Imputation**

$$y_{2,i} = \beta_0 + \beta_1 y_{1,i}$$



✓ Use of information from the observed data, more reliable results

✗ Over-estimation of model fit and variance weakening

Data Processing

Strategies for Handling Missing Data

Imputations Methods: **Stochastic Regression Imputation**

Same method as deterministic one but a stochastic term (e_i) is added to the regression formula to consider the spread of the data along the regression line

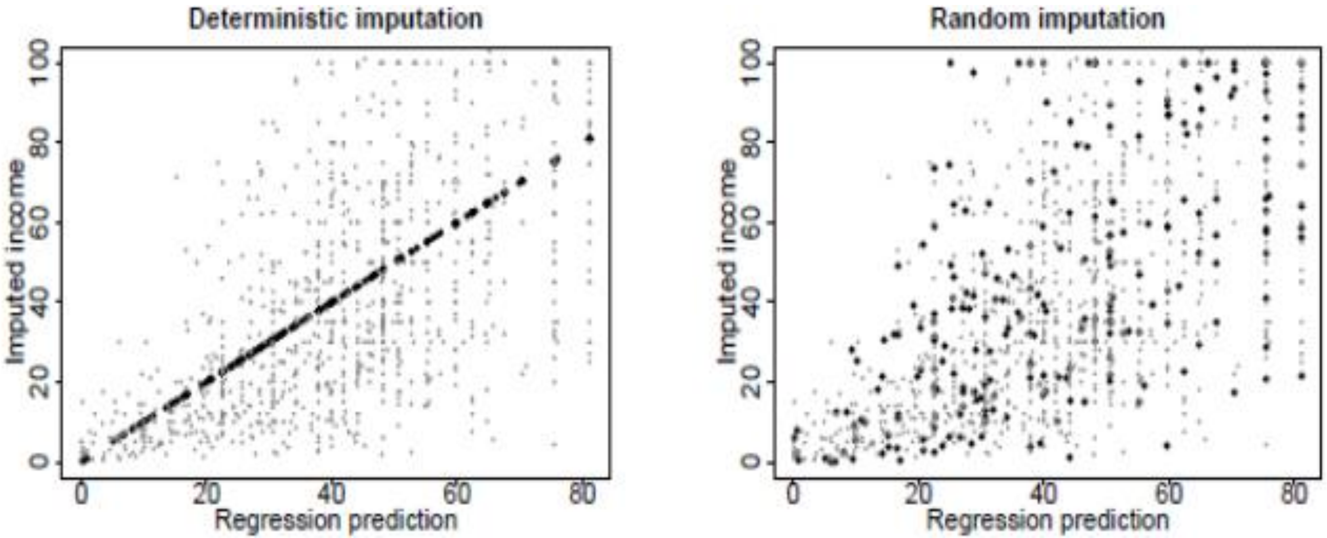
$$y_{2,i} = \beta_0 + \beta_1 y_{1,i} + e_i$$

Usually, e_i has a zero mean and variance equal to the error variance in regression

Data Processing

Strategies for Handling Missing Data

- **Imputations** Methods: **Regression**



Data Processing

Strategies for Handling Missing Data

Imputations Methods: **Multiple Imputation**

Repeated imputation procedure and result combination

i. Introduce random variation into the process of imputing missing values (stochastic method) and generate multiple datasets, each with slightly different values

ii. Analyse the datasets

iii. Combine results into a single set of parameters estimates, standard errors, statistics

✓ Decreased risk of bias, good estimates of

✗ More complex algorithm

Hands-On: Heart Disease Dataset

- **Missing Data**

https://tinyurl.com/3ywaw4wr