Co-financed by the Connecting Europe Facility of the European Union

Transforming Healthcare

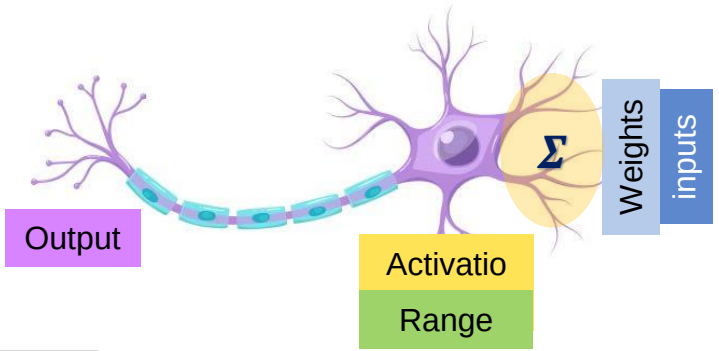
Data Science Process in Healthcare

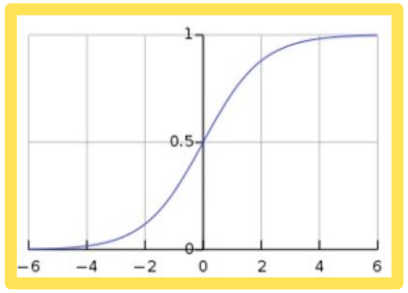
Bonfitto Giuseppe
bonfitto.giuseppe@hsr.it

Co-financed by the Connecting Europe Facility of the European Union

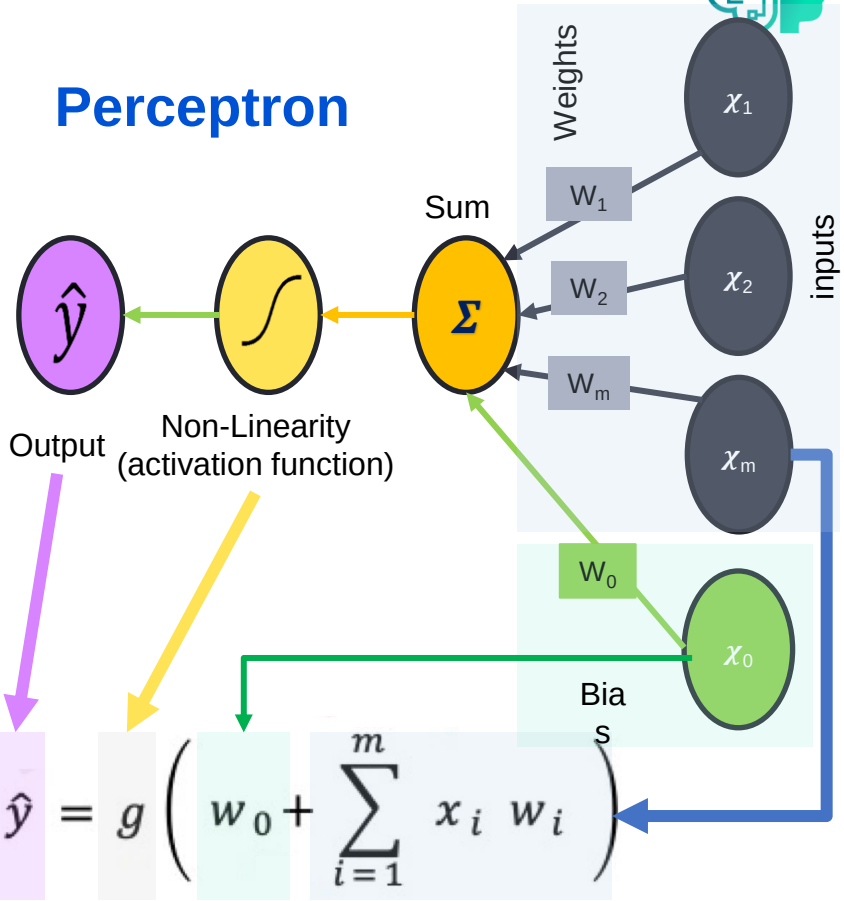
AI Modeling

Neural Networks





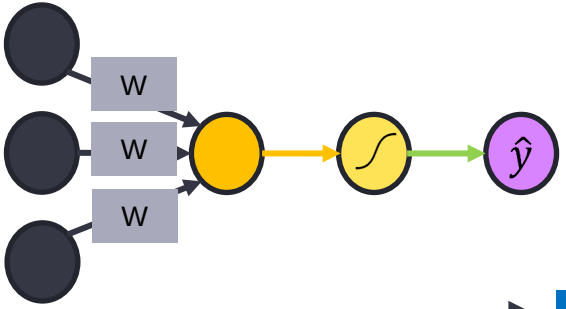
Perceptron



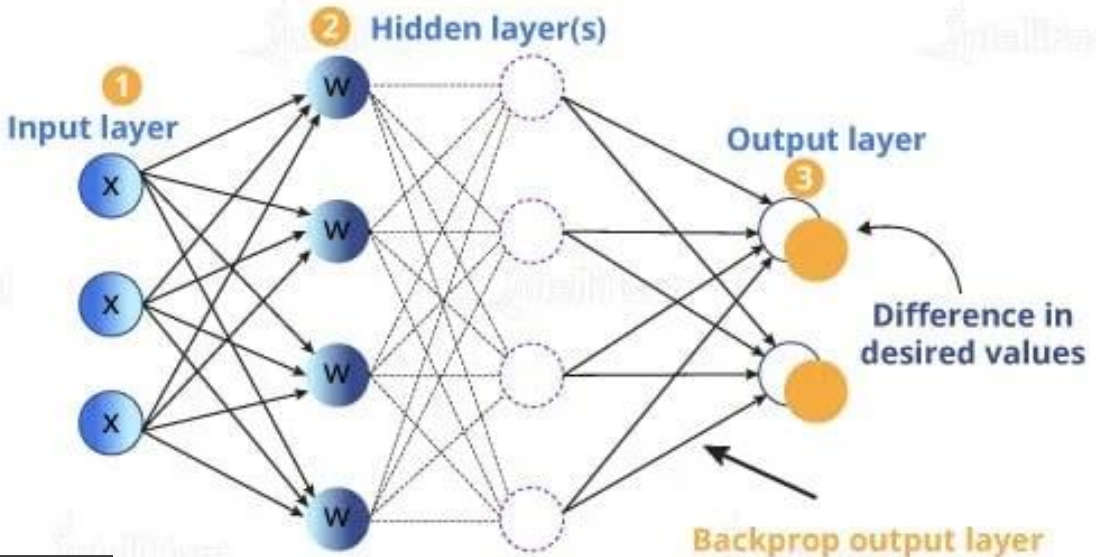
$$\hat{y} = g\left(w_0 + \sum_{i=1}^{m} x_i\, w_i\right)$$

AI Modeling

Neural Networks

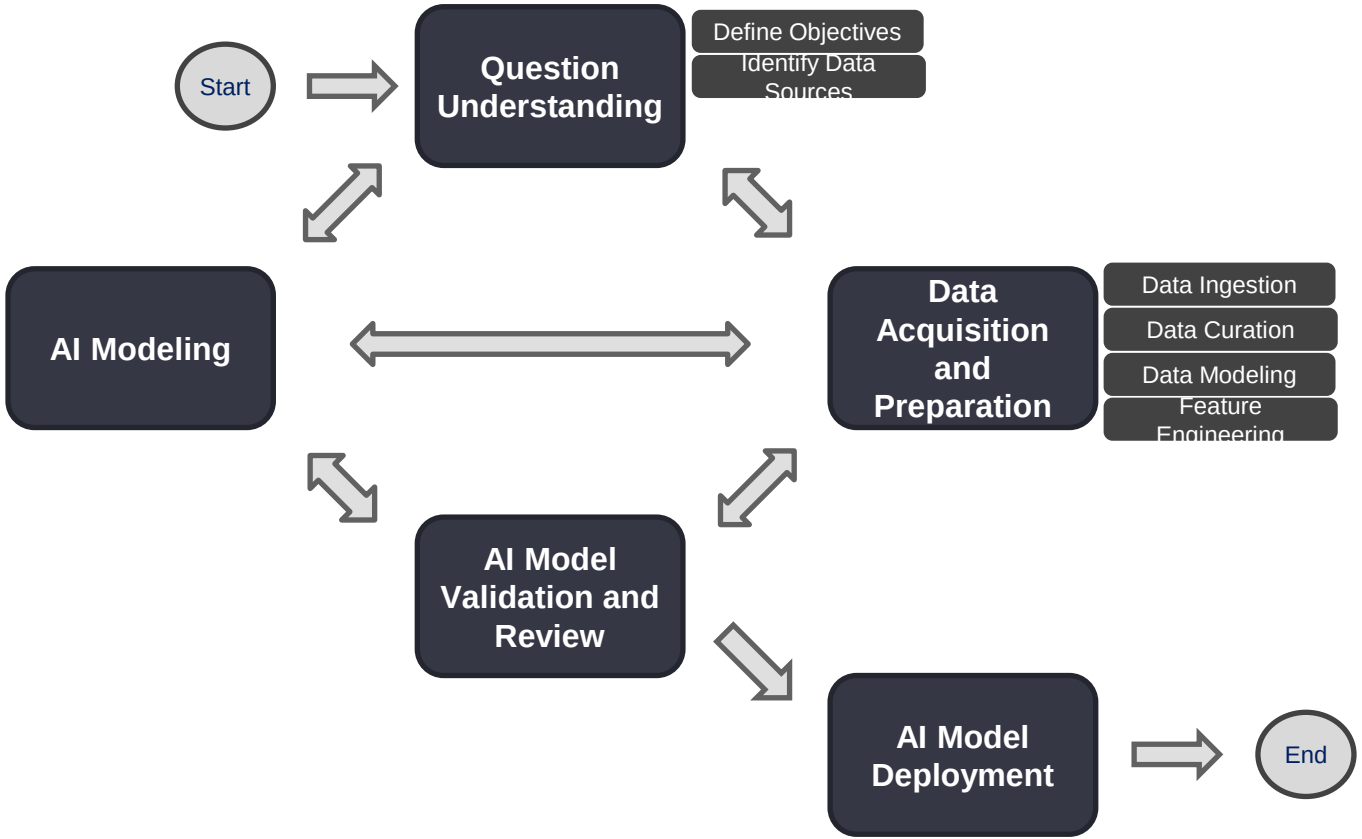


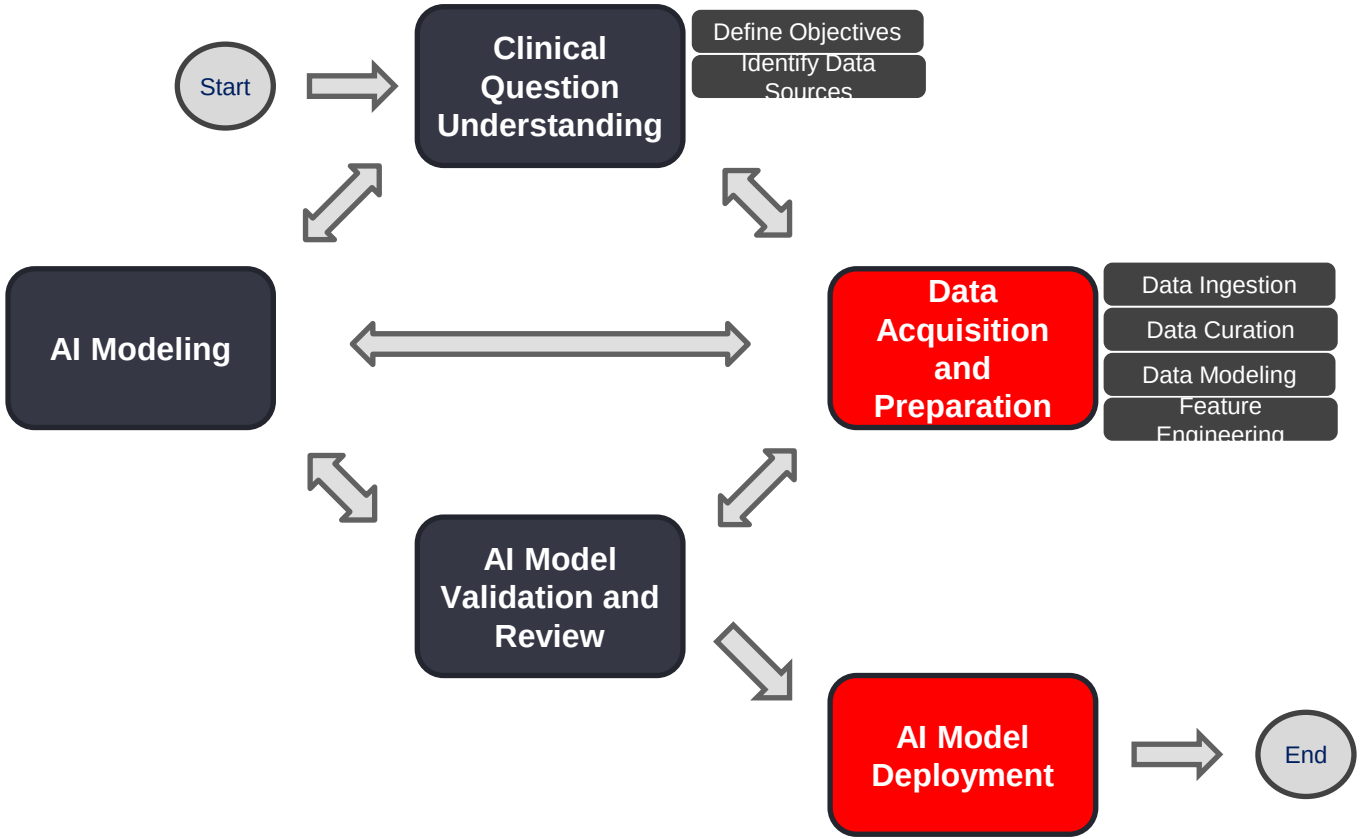
Forward propagation



Backpropagation

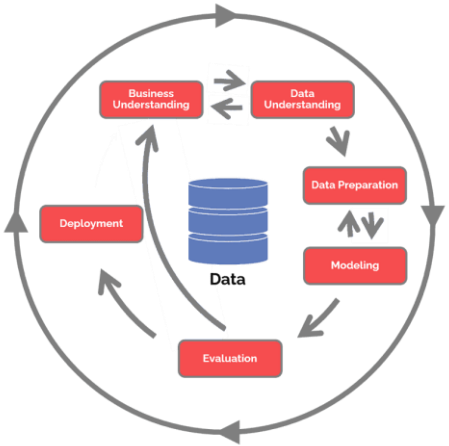
Slide Credit: IntelliPaat

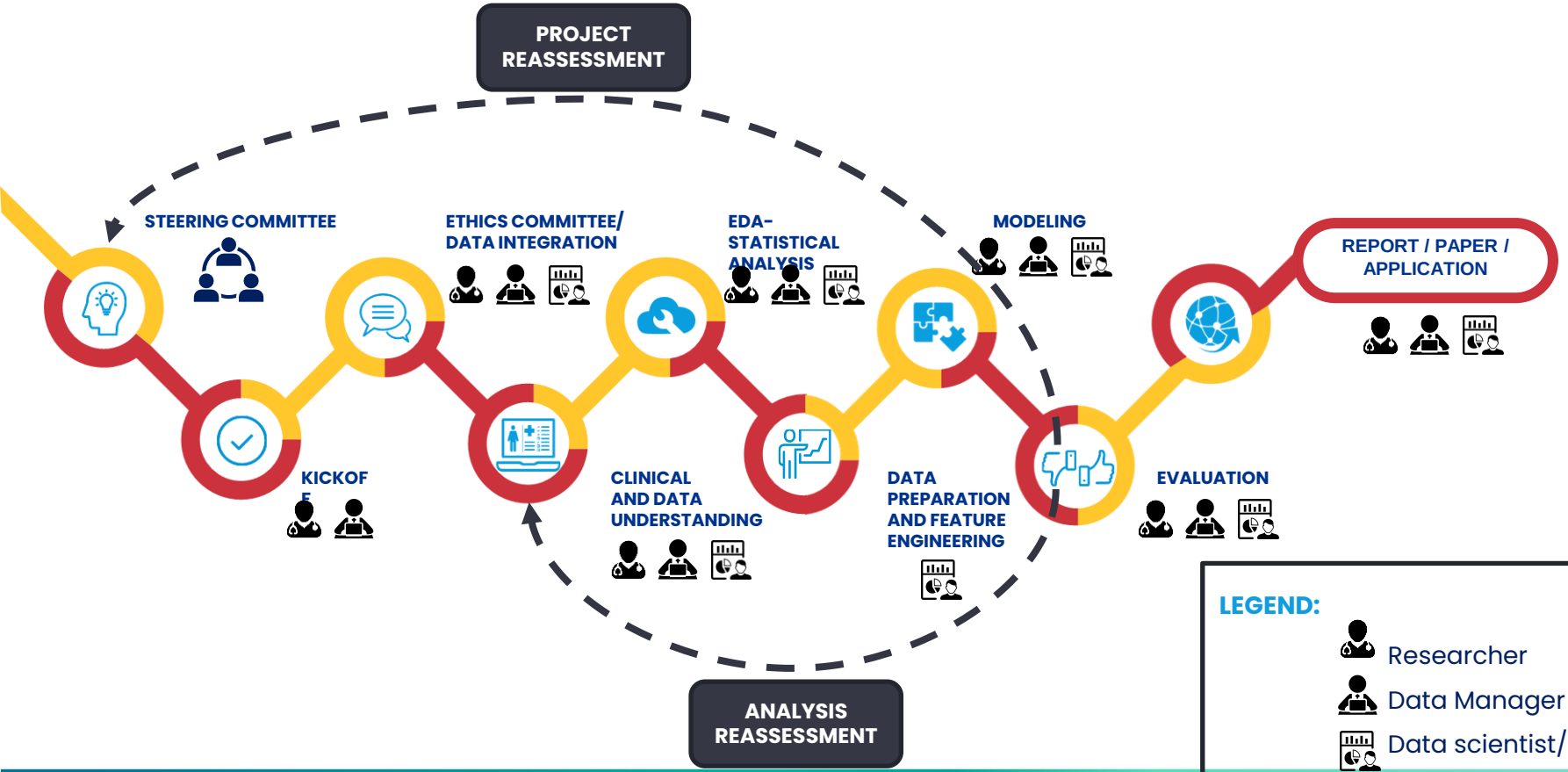

Start
Question Understanding
Define Objectives
Identify Data Sources
Data Acquisition and Preparation
Data Ingestion
Data Curation
Data Modeling
Feature Engineering
AI Modeling
AI Model Validation and Review
AI Model Deployment
End


Start
Clinical Question Understanding
Define Objectives
Identify Data Sources
Data Acquisition and Preparation
Data Ingestion
Data Curation
Data Modeling
Feature Engineering
AI Model Validation and Review
AI Modeling
AI Model Deployment
End

CRoss Industry Standard Process for Data Mining (CRISP-DM)

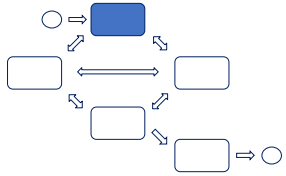
Clinical Problem Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Report and Paper
Collaboration with **clinicians** to **understand the context**, identify **goals** and **requirements**, and translate it into a **machine learning workflow**.	Collection, **exploratory analysis** (EDA), and **understanding** of **data** with the help of **clinicians** and **data managers**. Result: agreed and approved project plan and **sharing** of **analysis results** via ppt/dashboard..	Data **optimization** through **cleaning**, **encoding**, **feature engineering**, and eventually splitting the datasets for training and testing	Selection, training and optimization of **machine learning and deep learning** (ML and DL) and **traditional statistical algorithms**	Model testing, **performance evaluation**, validation of **clinical relevance** of results, and **review** of the whole **process**. If necessary, return to an earlier stage.	Preparation of a detailed **report** and, if necessary, a scientific **paper**. **Operationalization of the code** to ensure its reproducibility, future maintainability and prepare for possible deploy of the models.

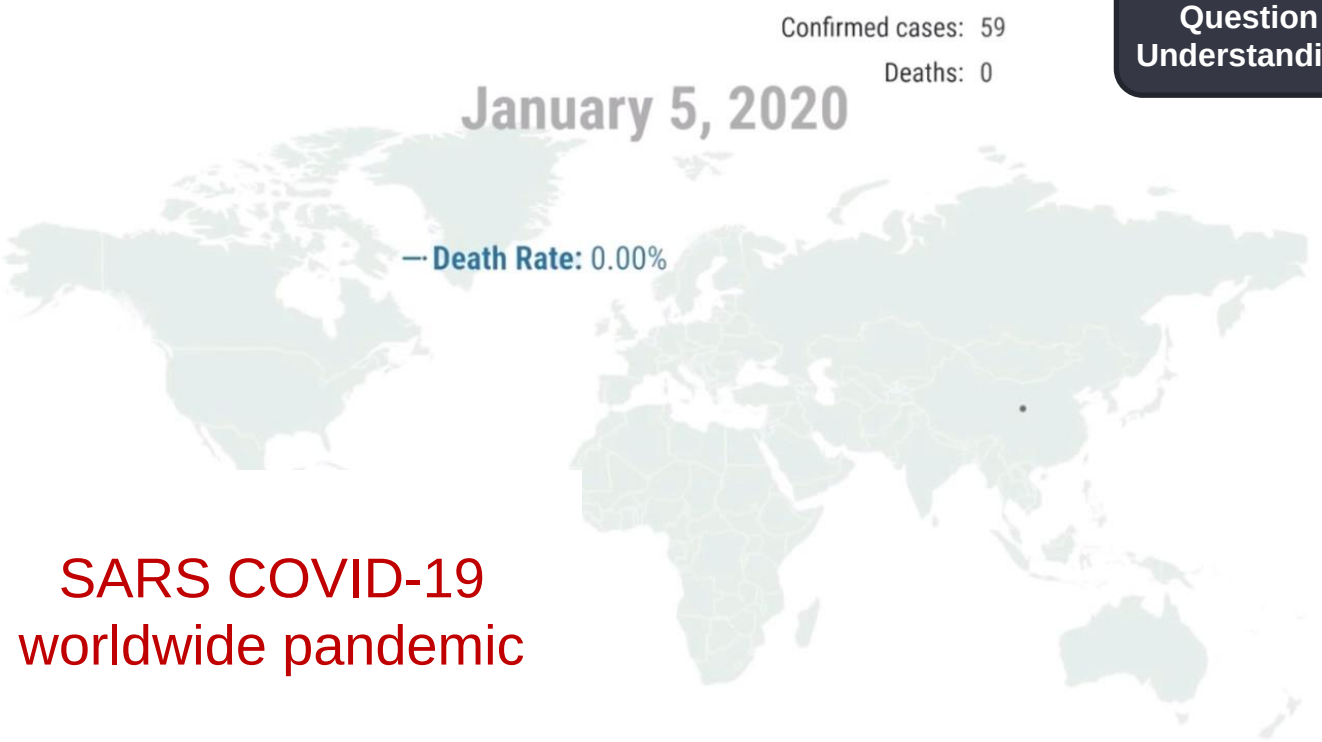



PROJECT REASSESSMENT
STEERING COMMITTEE
KICKOF F
ETHICS COMMITTEE/ DATA INTEGRATION
CLINICAL AND DATA UNDERSTANDING
EDA- STATISTICAL ANALYSIS
DATA PREPARATION AND FEATURE ENGINEERING
MODELING
EVALUATION
REPORT / PAPER / APPLICATION
ANALYSIS REASSESSMENT
LEGEND:
Researcher
Data Manager
Data scientist/analyst

Co-financed by the Connecting Europe Facility of the European Union

Clinical Question Understanding



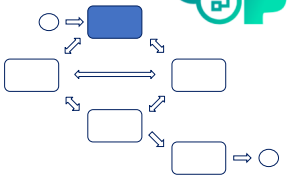


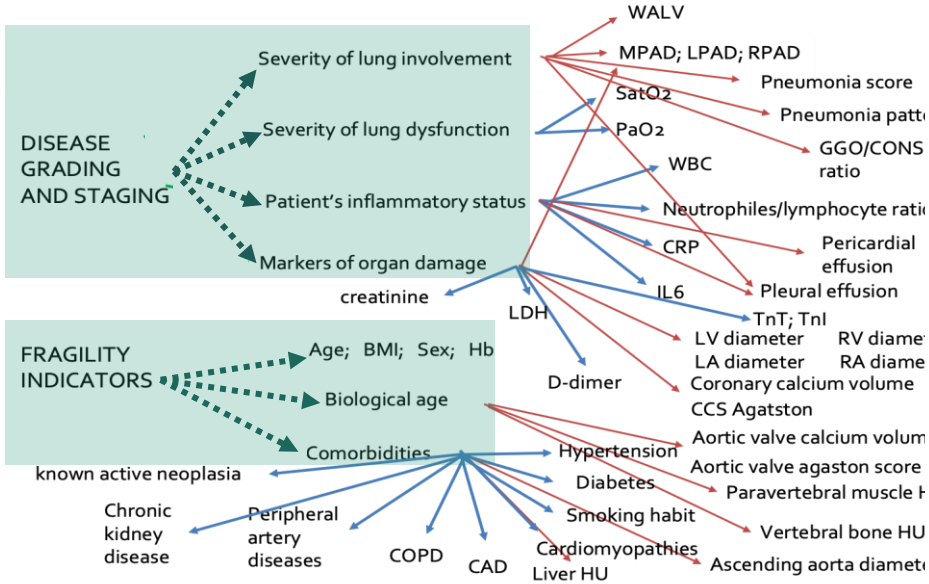
SARS COVID-19
worldwide pandemic

Clinical Question Understanding

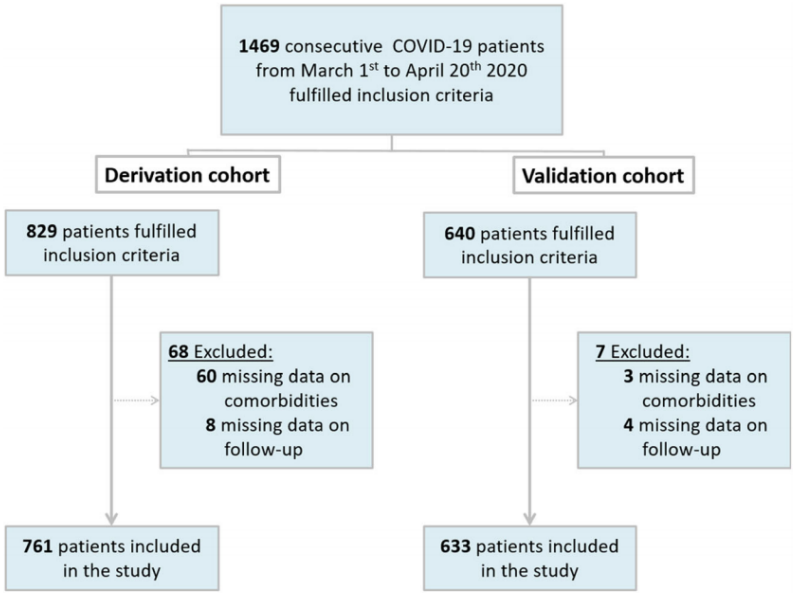
Define Objectives

Identify Data Sources





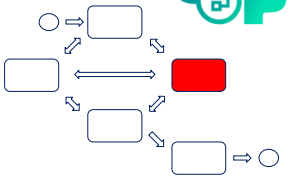
sCORE network



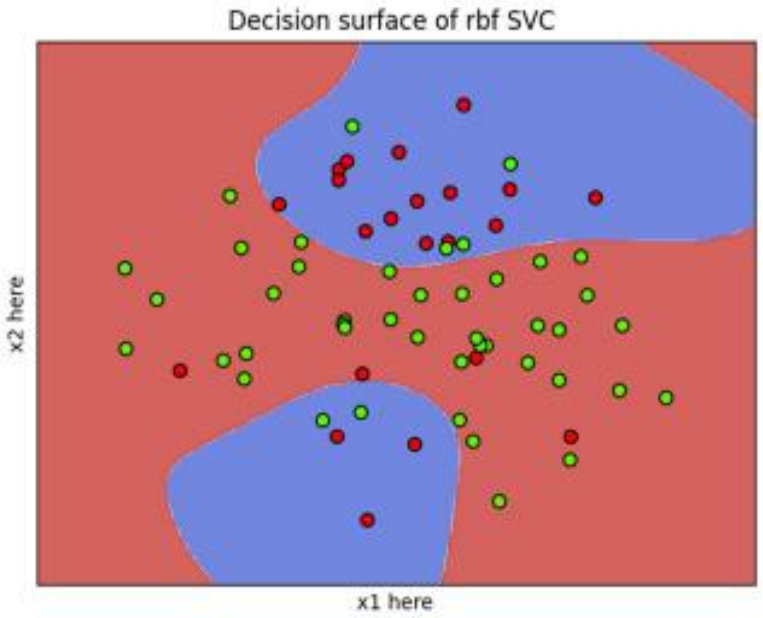
Thoracic CT parameters

Clinical and laboratory parameters

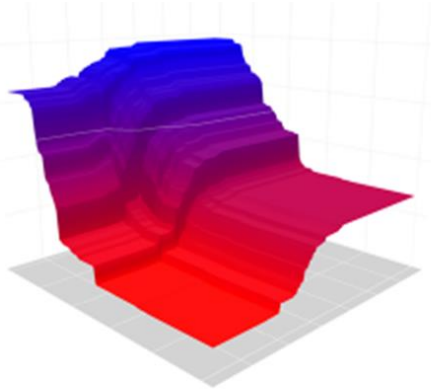
Co-financed by the Connecting Europe Facility of the European Union



Data Acquisition and Preparation



SVM model after dimension reduction



ADA Boost 3D after dimension reduction

Other analysis:

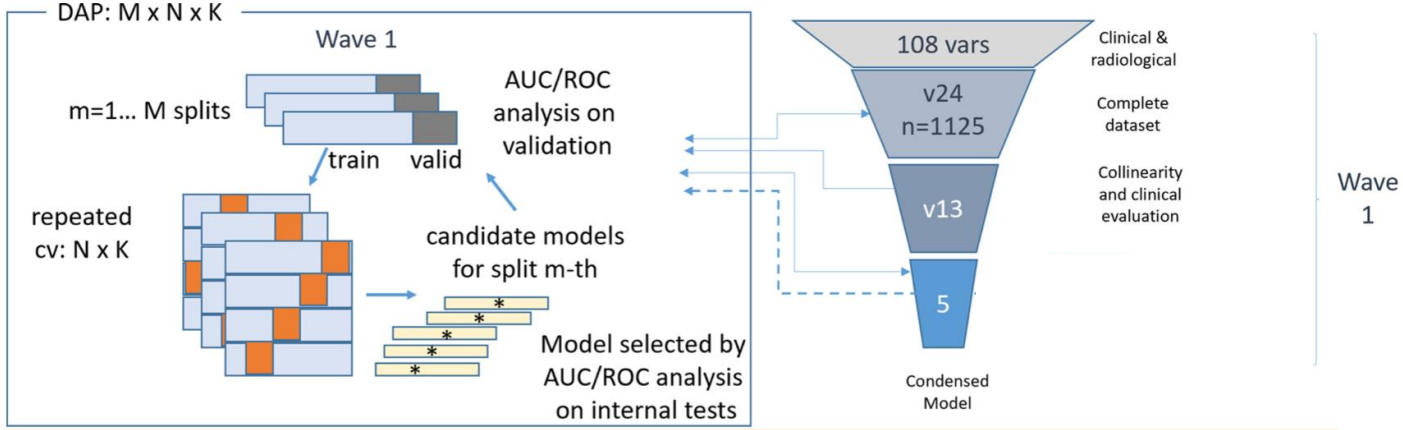
Most informative content: analysis to find features with most discrimination power

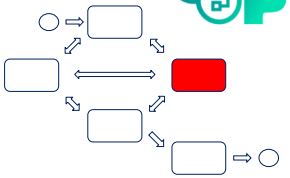
Univariate: analysis to find correlation between single feature and class

Covariate: analysis to find correlation between group of features and class

Misclassification: analysis to find factors that encourage wrong class attribution

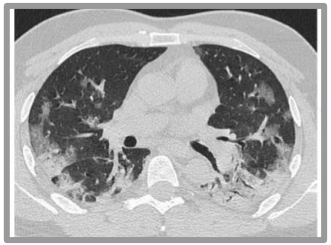
Co-financed by the Connecting Europe Facility of the European Union



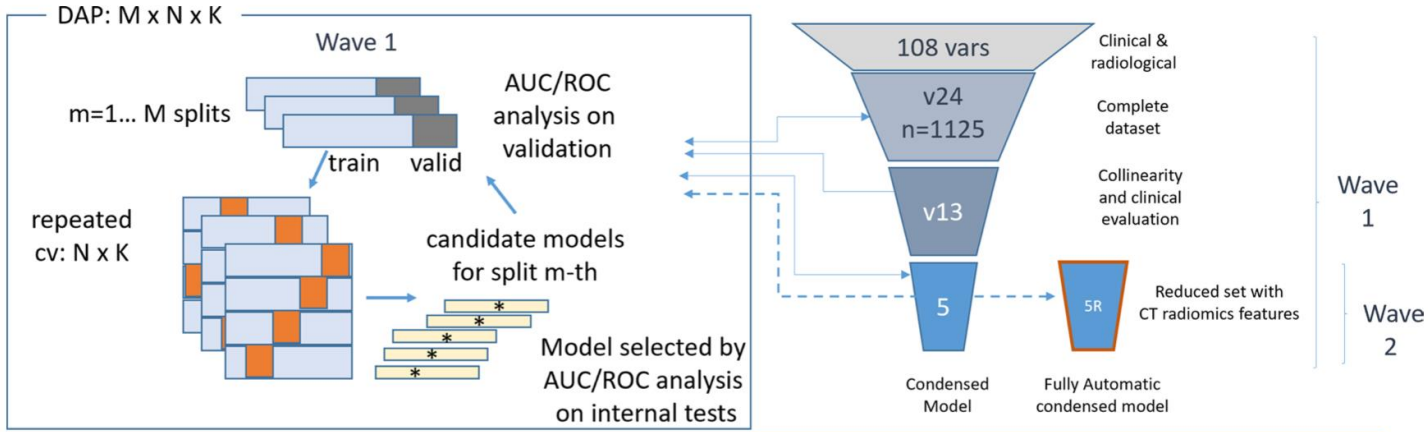


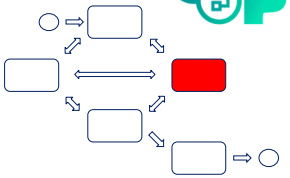
Data Acquisition and Preparation

Bayesian Logistic Regression: Classification using manual CT extraction		
Target outcome	**Number of selected features**	**Result**
Survival	24	AUC = 0.839±0.009
Survival	13	AUC=0.840±0.0093
Survival	5	AUC = 0.834 ± 0.007



Co-financed by the Connecting Europe Facility of the European Union





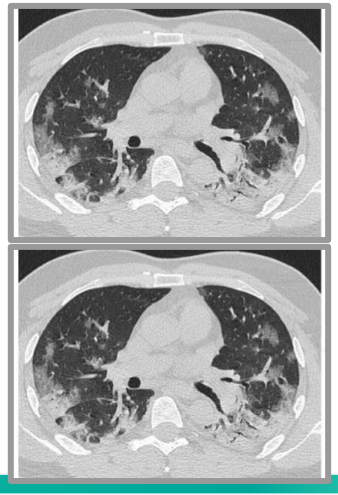
Data Acquisition and Preparation

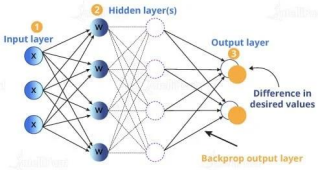
Bayesian Logistic Regression: Classification using manual CT extraction

Target outcome	Number of selected features	Result
Survival	24	AUC = 0.839±0.009
Survival	13	AUC=0.840±0.0093
Survival	5	AUC = 0.834 ± 0.007

Bayesian Logistic Regression: Classification using fully automatic CT extraction

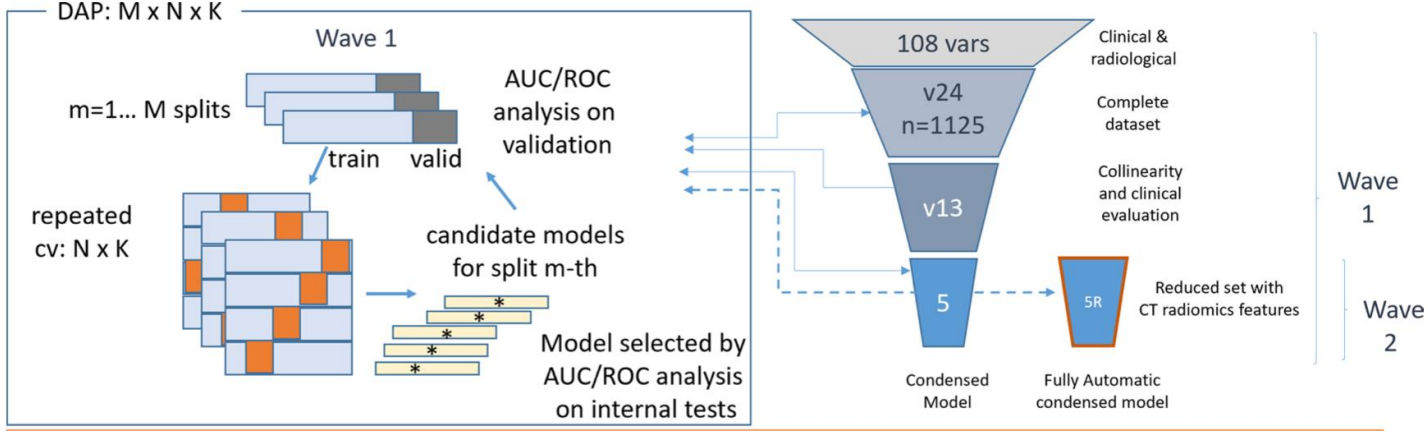
Target outcome	Number of selected features	Result
Survival	5	AUC=0.842 (DeLong 95% CI: 0.816-0.867)





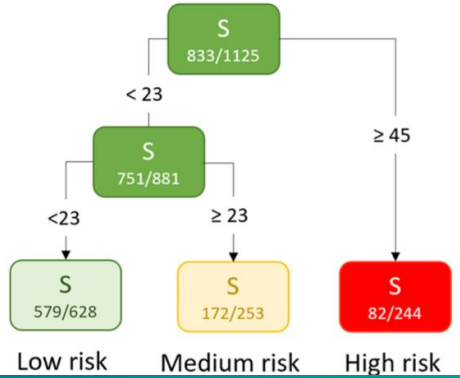
eXplainable Artificial Intelligence in healthcare Management
2020-EU-IA-0098

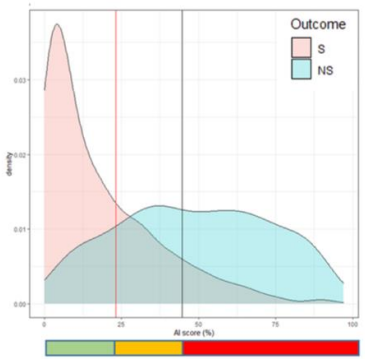
Co-financed by the Connecting I Facility of the European Union

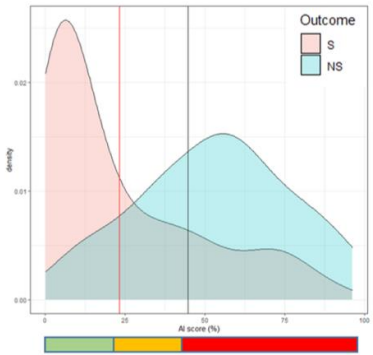


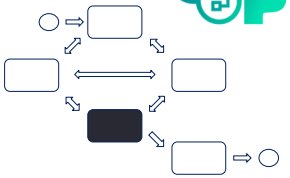
Model selection and validation by DAP on Wave 1: vars24 → vars13 → var5 → 5R

Tree-based binning in 3 risk classes → AI-SCoRE Wave1 → AI-SCoRE Wave2



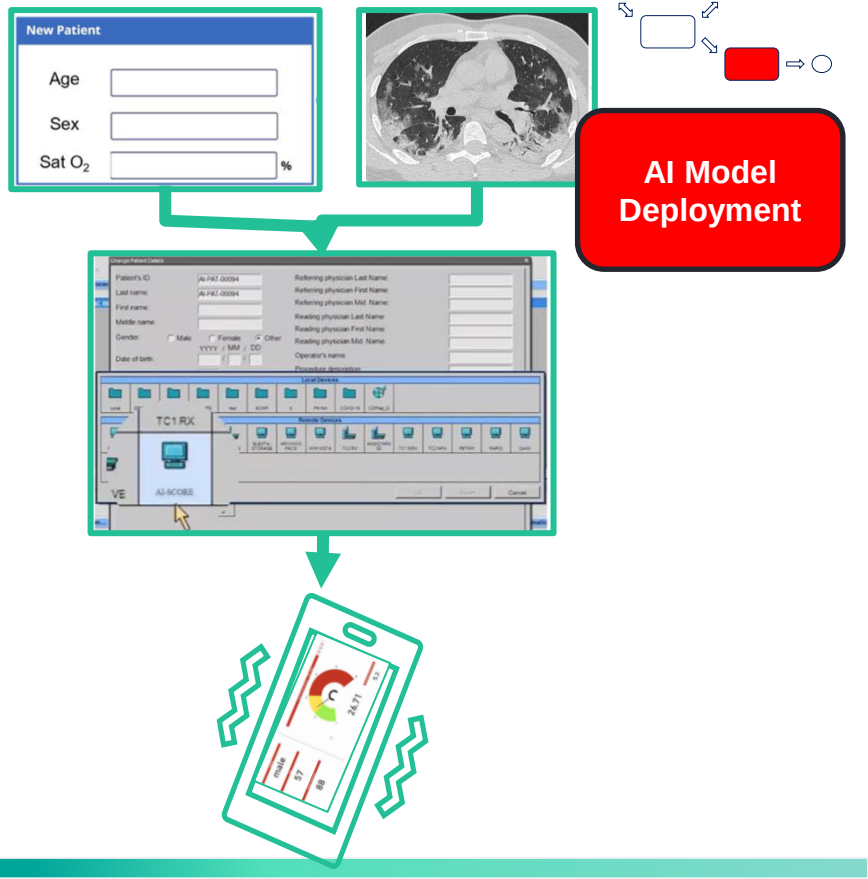




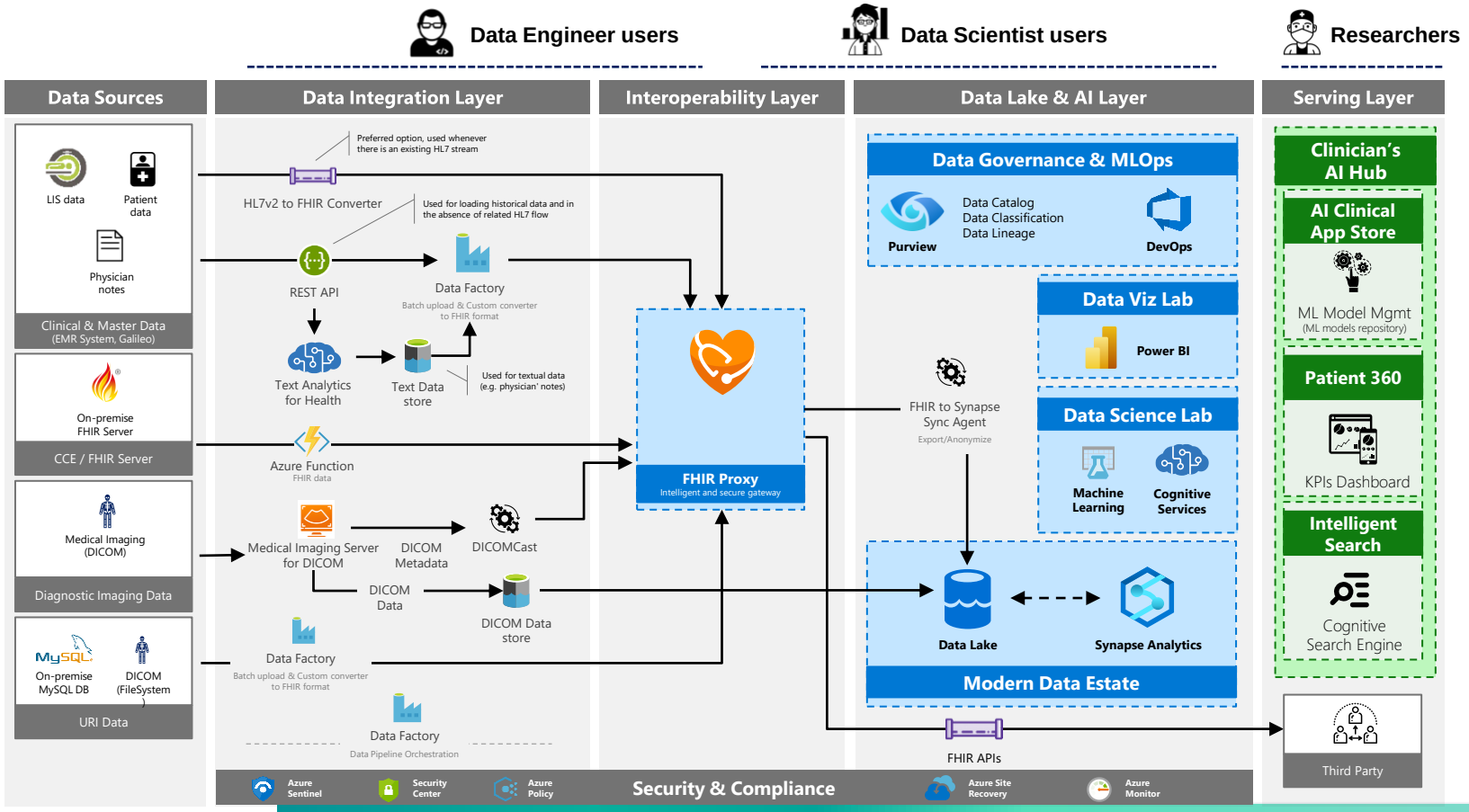


AI Model Validation and Review


Number of Patient
1673
Number of prediction available
7
% Death Male over Female
67,77%
Number of Patient by Area
Death Rate
Number of Patient by Hospital and Gender
Number of Patient by Gender
Death/Survival Distribution by Date
Death/Survival Distribution by Age
Pagine
Patient AI Score
Patient Details
AiScore Patients List
Patient identifier
Hospital
Gender
Age
AI SCORE
BOL
male
72
(Vuoto)
Patient Observations
Observation Comparison
Encounter Start
Encounter End
Hospitalization Days
2020-03-17
2020-03-30
14

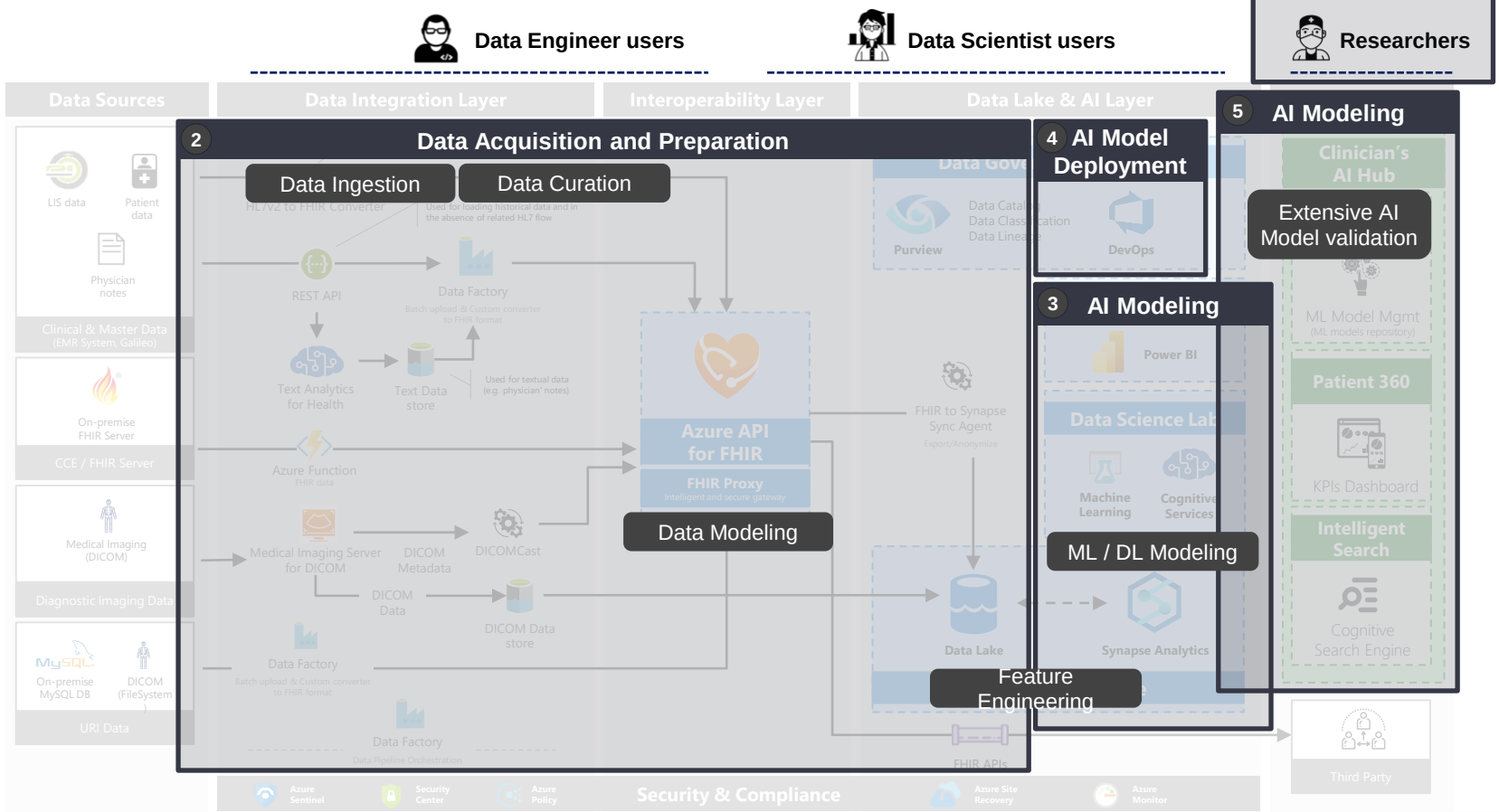

New Patient
Age
Sex
Sat O_2
%
AI Model Deployment

Microsoft Intelligent Data Platform, Azure Health Data Services, Azure Cognitive Services & Azure Machine Learning

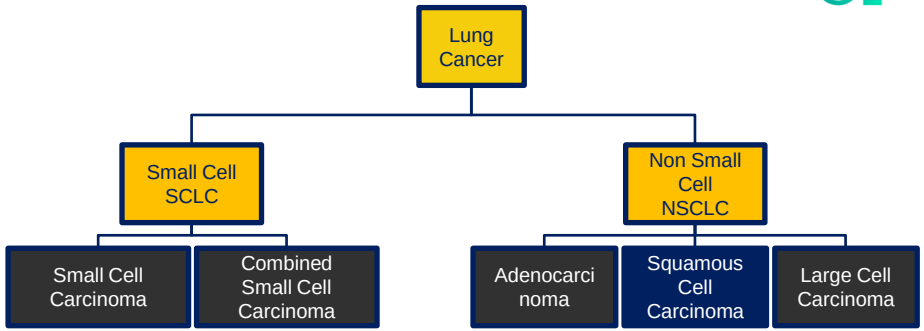


Co-financed by the Connecting Europe Facility of the European Union

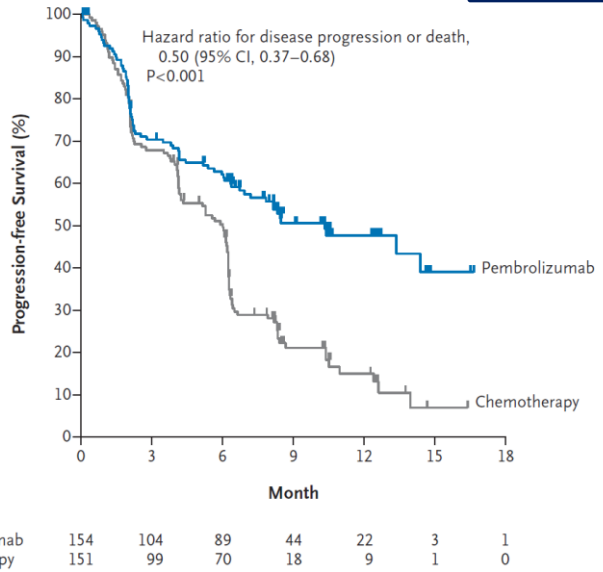
eXplainable Artificial Intelligence in healthcare Management
2020-EU-IA-0098

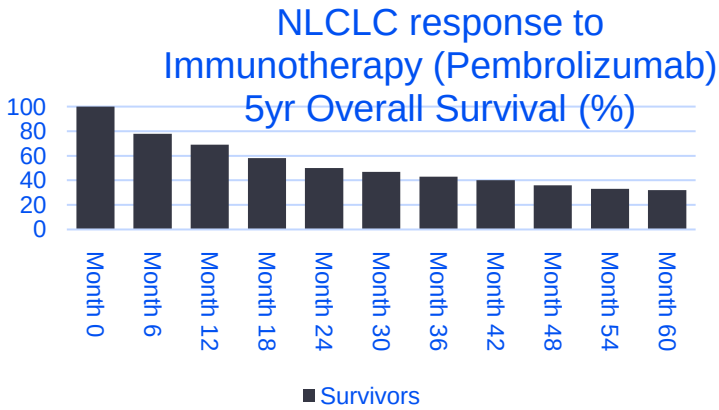


Co-financed by the Connecting Europe Facility of the European Union



PD-L1>50%
No EGFR mut
--
Pembrolizumab (first line)





Data Source: Reck et al New England J Medicine 2016,375:1823

Co-financed by the Connecting Europe Facility of the European Union

NSCLC – PDL1+ Retrospective Study

OUTCOME
Short survival (3 mo) or stable progressive response

71 variables

46 Clinical
(Symptomps, Comorbidity, Drugs, Follow-up)

12 Laboratory
(Leuco, Neutro %, Lympho %, HB, N/L RATIO, PLT, PCR, AST, ALT, BILI)

13 Pathology
(PDL1, ALK, ROS1, BRAF, KRAS, EGFR, RET, MET, NTRK)

EXCLUSION CRITERIA
- 1st Thorax CT (contrast) not available
- Not non-negligible disease

INCLUSION CRITERIA
- Metastatic NSCLC Stage IV
- Pembrolizumab (1st line therapy)
- PDL1>50%

106 Patients

54 non PD

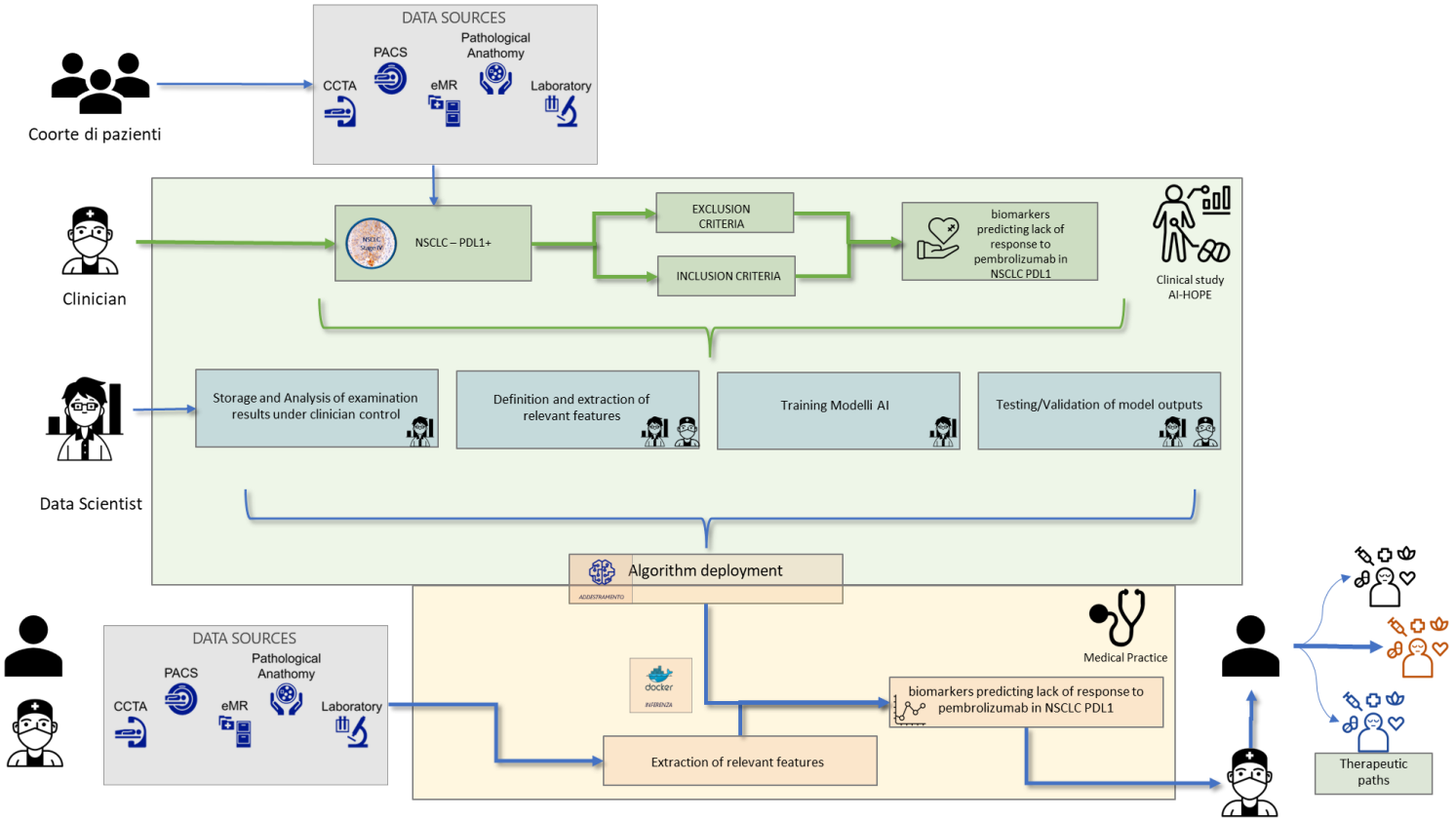
52 PD

Clinical need: Identify specific biomarkers predicting lack of response to pembrolizumab in NSCLC PDL1+

All data presented are unpublished and confidential

Co-financed by the Connecting Europe Facility of the European Union

eXplainable Artificial Intelligence in healthcare Management
2020-EU-IA-0098



Co-financed by the Connecting Europe Facility of the European Union

Data Sources

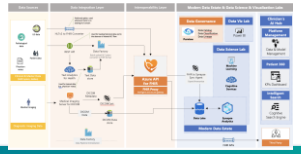
- LIS data
- Pathological data
- Patient data
- Physician notes

Clinical & Master Data (EMR System, Galileo)

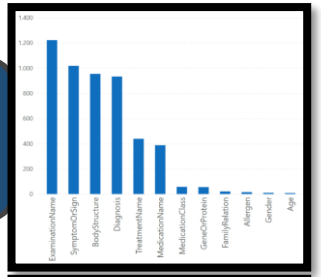
Medical Imaging

Diagnostic Imaging Data

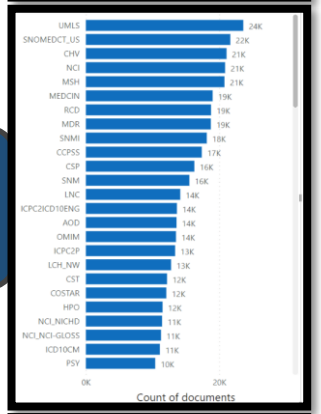
- detect **words** and **phrases** mentioned in **unstructured text** as **entities** that can be associated with **semantic** types in the **healthcare** and biomedical domain
- connect entities to **medical ontologies** and domain-specific coding systems (example UMLS, SNOMED, LOINC)
- identifies **meaningful links between concepts** mentioned in text



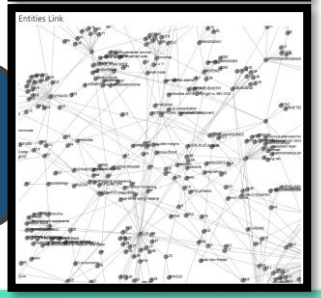
ENTITIES



MEDICAL ONTOLOGIES



LINKS



Co-financed by the Connecting Europe Facility of the European Union

eXplainable Artificial Intelligence in healthcare Ma
2020-EU-IA-0098

Clinician's AI Hub

Platform Management

Data & Model Management

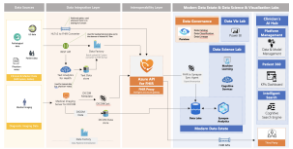
Patient 360

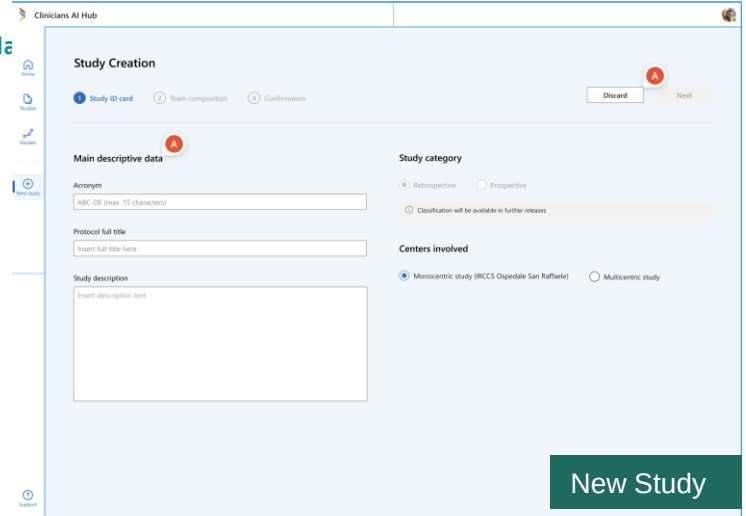
KPIs Dashboard

Intelligent Search

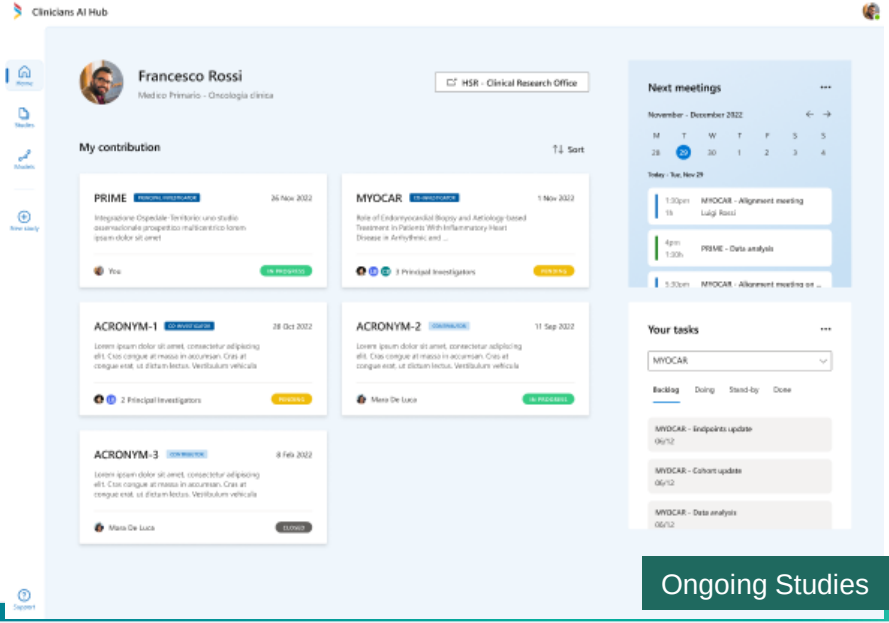
Cognitive Search Engine

Third Party

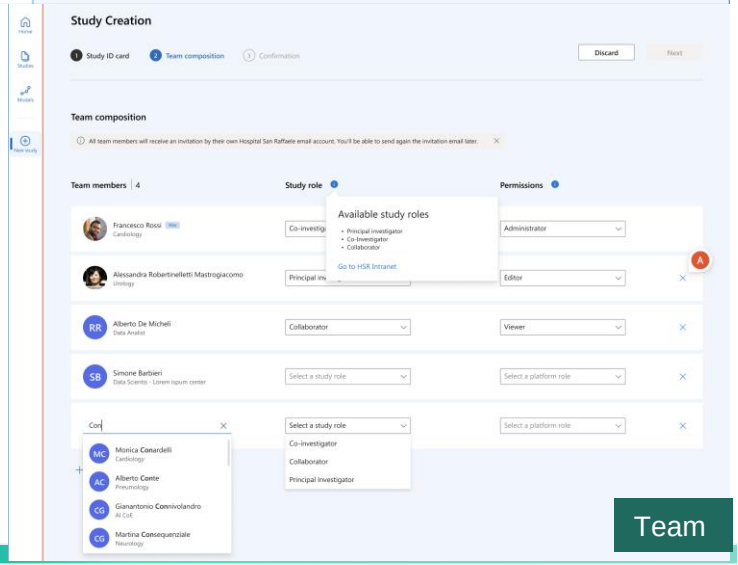




New Study



Ongoing Studies



Team

Co-financed by the Connecting Europe Facility of the European Union

eXplainable Artificial Intelligence in health
2020-EU-IA-0098

Clinician's AI Hub

Platform Management

Data & Model Management

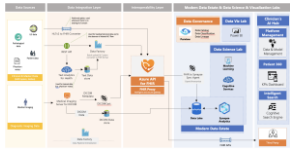
Patient 360

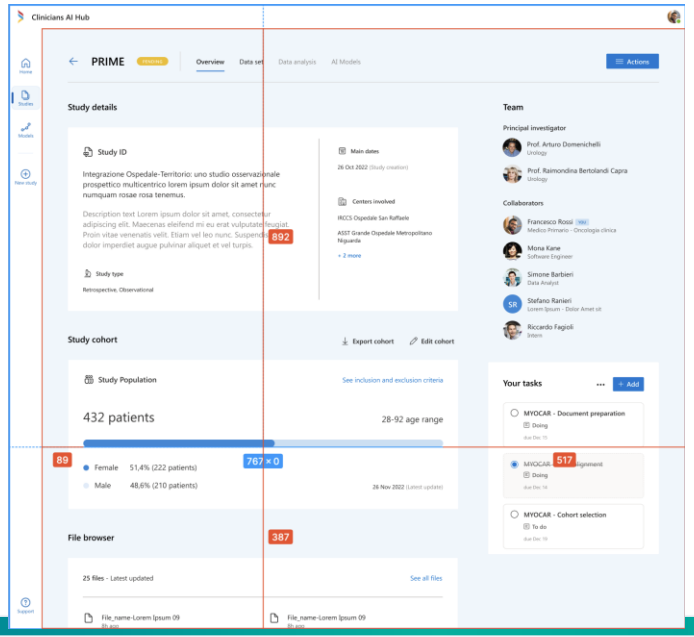
KPIs Dashboard

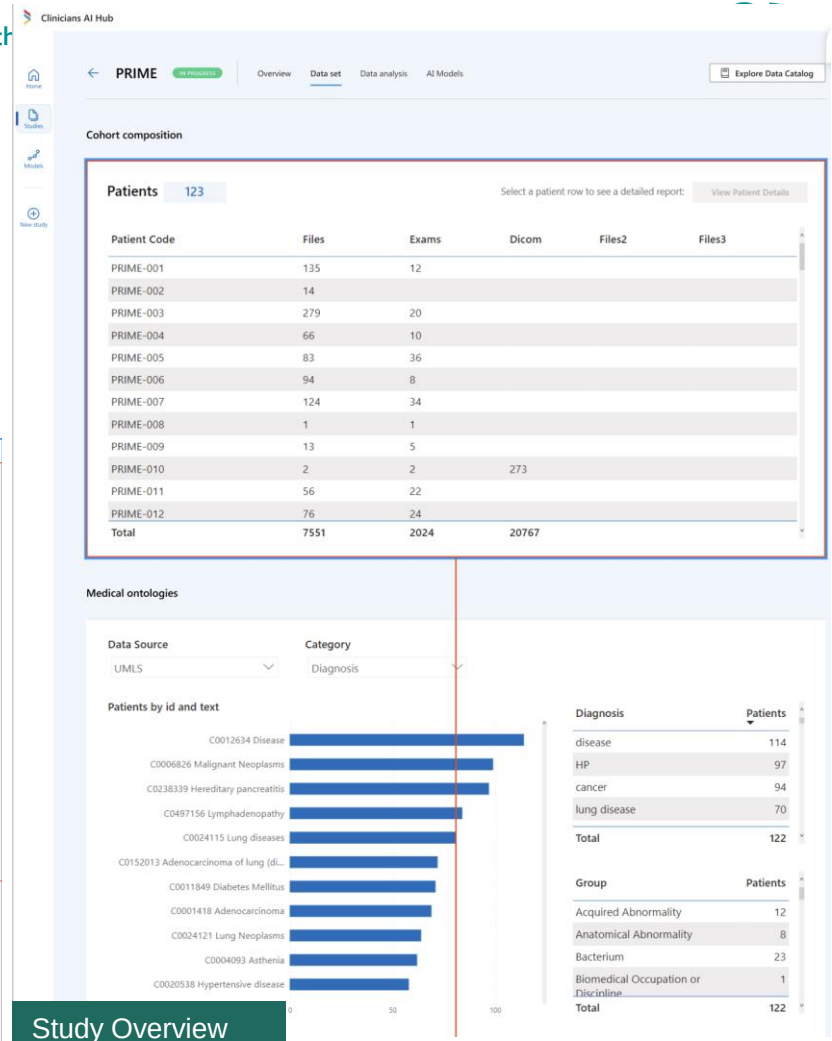
Intelligent Search

Cognitive Search Engine

Third Party







Study Overview