Deep Learning and Explainability for Multimodal Medical Data

Miriam Cobo Cano
PhD student at Institute of Physics of Cantabria (CSIC-UC)
Wilson Silva, Lara Lloret Iglesias
cobocano@ifca.unican.es

ABSTRACT

Screening with low-dose CT (LDCT) has the potential to **increase** the **proportion of lung cancer patients diagnosed** with **early-stage disease** who can be offered treatment with curative intent. This research aims to provide **explainable deep learning** (DL) algorithms in lung cancer screening to promote early detection integrating **multimodal data** from circulating proteome in pre-diagnostic plasma samples, LDCT scans, and clinical variables.

OBJETIVES

- Leverage **multimodal data in explainable DL** algorithms to develop a **nodule malignancy risk score model** including imaging features and protein markers.
- Optimize clinical **management** of **indeterminate pulmonary nodules** detected on LDCT screening for lung cancer.
- Correlate radiological **imaging features** with novel **protein biomarkers** to improve **early detection** of **lung cancer**.

METHODS

The **challenge** in early diagnosis of lung cancer is that **lung nodules** are found in approx. **25% of screening participants**, but **only** a **small fraction** are **malignant**.

Our dataset **LungAmbition** is a single center dataset from Clínica Universidad de Navarra (Spain):

- Protein data acquired using **Olink proteomics platform**. **Circulating proteome** in prediagnostic plasma samples was quantified using the proximity extension assay (PEA).
- LDCT scans include **the segmentation of the nodule** subsequently diagnosed as malignant in a follow-up study.
- Several **clinical variables** are available.

The groups of patients are **matched by sex and age**:

- 76 patients with malignant nodules (lung cancer).
- 76 patients with benign nodules.
- 7 false positive patients.
- 76 control patients (no nodules).

Prediagnosis plasmas samples and LDCT scans are collected within **5 years** prior to lung cancer diagnosis.
This dataset is part of to the **INTEGRAL program** (Integrative Analysis of Lung Cancer Etiology and Risk). The project will follow a **FAIR methodology.**

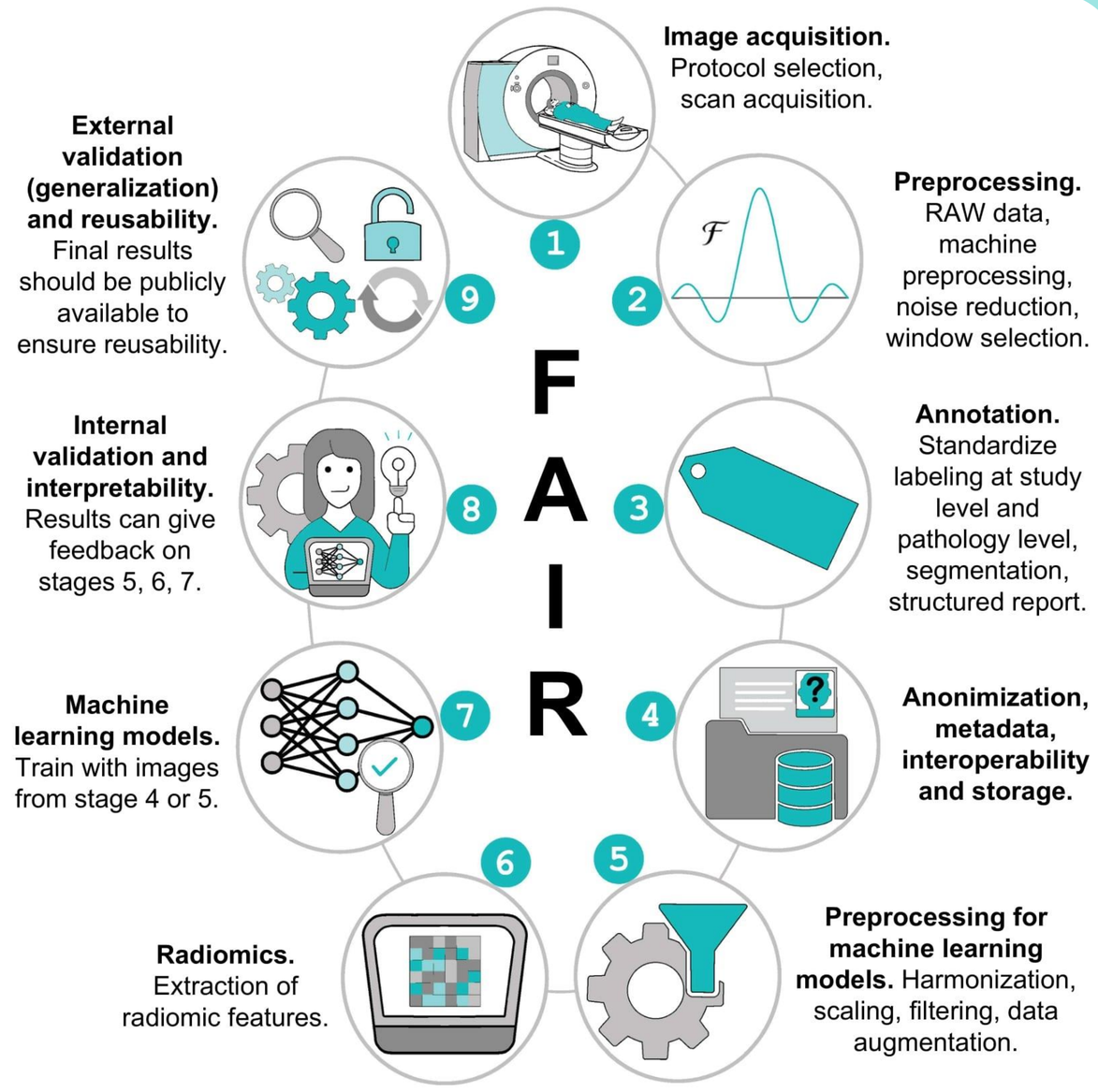

Figure 1. Workflow to develop FAIR principles in medical imaging data [1].

RESULTS

Recent study [2] from INTEGRAL consortium on **circulating proteome** from prediagnosis plasma samples, including 4 cohorts (Canada, Spain –our cohort-, USA, UK):

- **36 potentially informative circulating protein markers** identified differentiating malignant from benign nodules.
- **10** markers were particularly **relevant for imminent cancer diagnoses** within 1 year.
- Identification of **biological pathways beyond inflammatory-related** or **cancer-related proteins**.
- **Protein** marker **expression levels differ** with the **presence of pulmonary nodules**, supporting the use of patients with benign nodules as the main comparison group, and healthy controls only to evaluate background expression level.

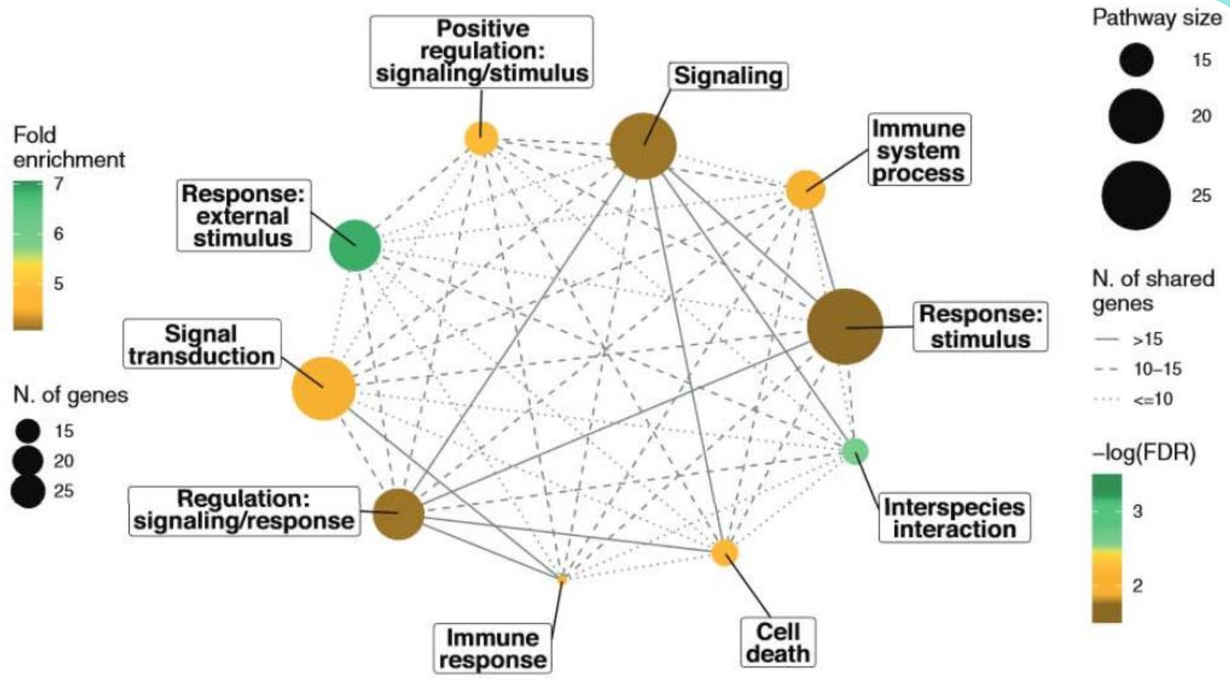

Figure 2. Network analysis of significantly enriched pathways based on the 36 informative markers [2].

CONCLUSIONS

- The findings will contribute to **create a customized panel** to **absolutely quantify targeted protein markers** for lung cancer risk and pulmonary nodule malignancy assessment.
- **Explainable multimodal DL algorithms** shed light on decision-making understanding, enhancing the confidence of clinicians and patients.
- The **future of personalization** in cancer research calls for collaborative efforts in highly interdisciplinary research.

REFERENCES

1. Cobo, Miriam, et al. "Enhancing radiomics and Deep Learning systems through the standardization of medical imaging workflows." Scientific Data 10.1 (2023): 732.
2. Khodayari Moez, Elham, et al. "Circulating proteome for pulmonary nodule malignancy." JNCI: Journal of the National Cancer Institute 115.9 (2023): 1060-1070.
3. Robbins, Hilary A., et al. "Design and methodological considerations for biomarker discovery and validation in the Integrative Analysis of Lung Cancer Etiology and Risk (INTEGRAL) Program." Annals of epidemiology 77 (2023): 1-12.