UNIVERSITÀ DEGLI STUDI DI BRESCIA

Artificial Intelligence in Medicine and Innovation in Clinical Research and Methodology
ARTIFICIAL INTELLIGENCE IN MEDICINE
XXXVIII CYCLE

Coordinatore: Prof. Domenico Russo

AI tools for enhanced follow-up of patients with low-back pain

PhD Student: **Alemu Sisay Nigru**

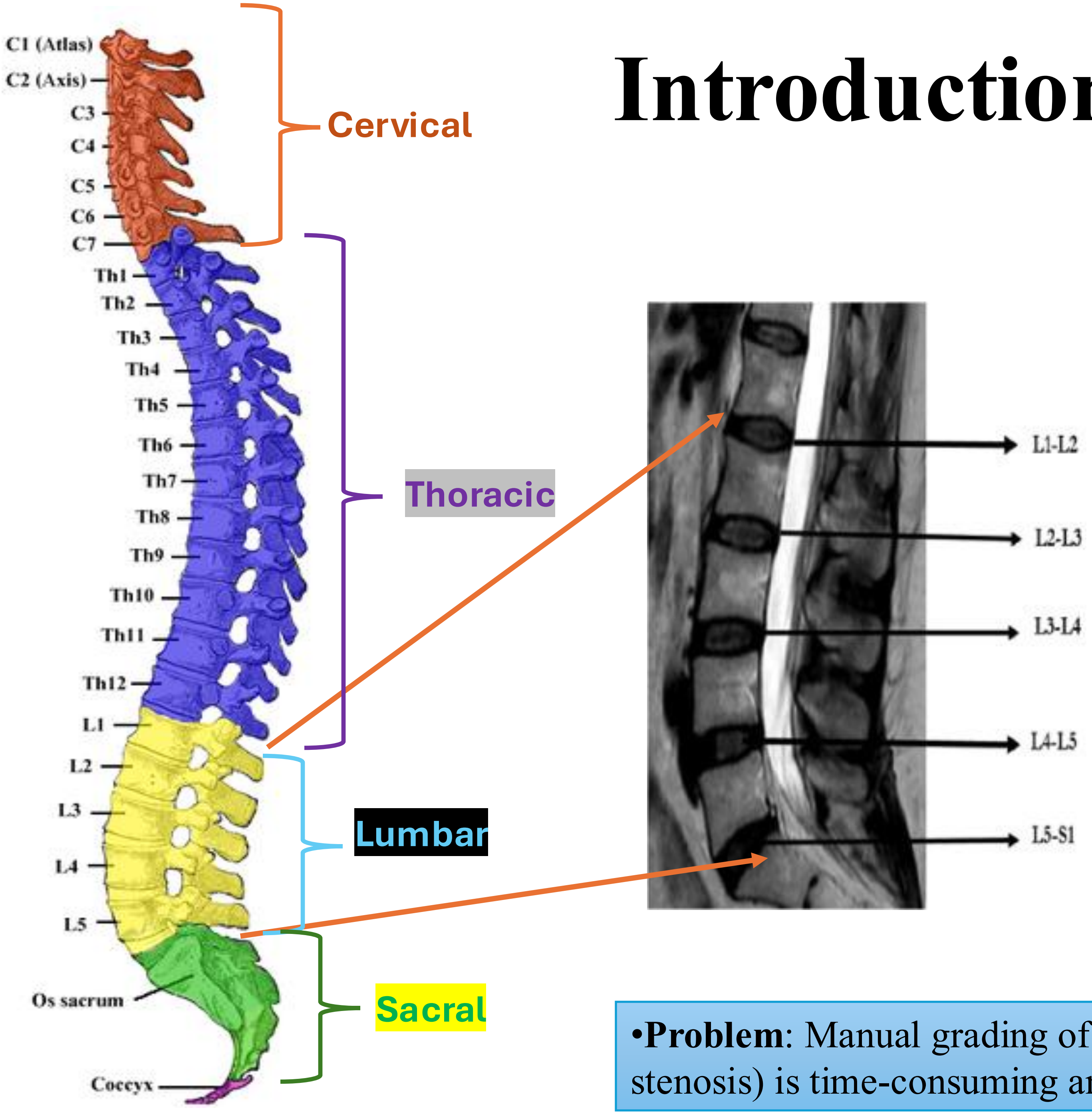
Supervisor: **Prof. Riccardo Leonardi**

May 7/2025

alemu.nigru@unibs.it

XRAYSERVICE NYU TANDON CDHI Center for Digital Health & Implementation Science

Introduction



Lumbar disc herniation is one of the most common intervertebral disc diseases (IDD), resulting in limited movement and unbearable pain levels.

Lumbar discs are small joints that lie between each two vertebrae (L1-L2, L2-L3, L3-L4, L4-L5 and L5-S1).

•**Problem**: Manual grading of disc pathologies (e.g., Pfirrmann grades, stenosis) is time-consuming and prone to inter-radiologist variability.

SOTA

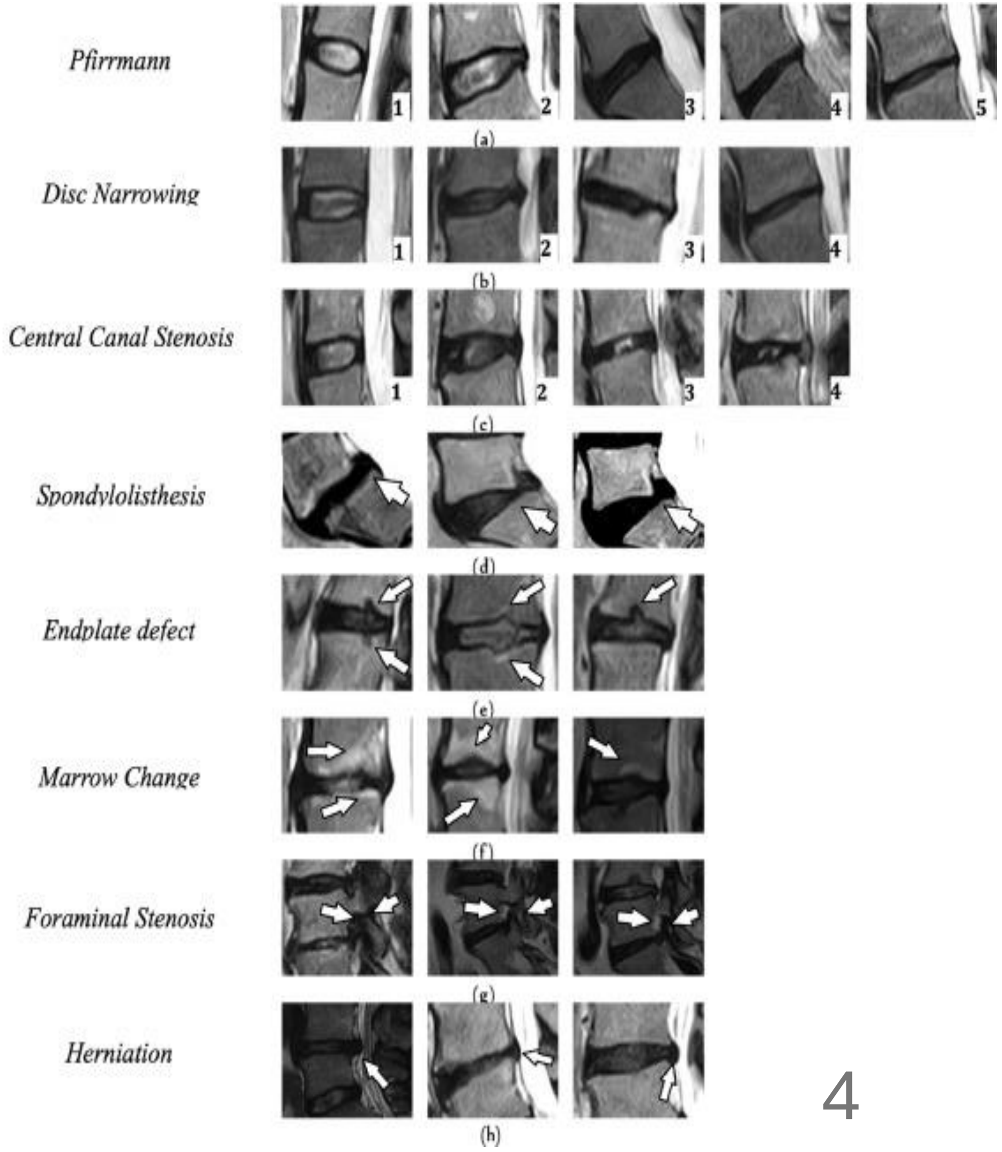
Background & Objective

- **Objective:** Validate **SpineNetV2**, an open-access DL model, for automated detection/grading of disc pathologies across diverse datasets.

- **Necessity of External Validation:** To ensure the ML model’s generalizability, robustness and real-world applicability, it is essential to evaluate its performance on unseen data through external data, and by external experts.

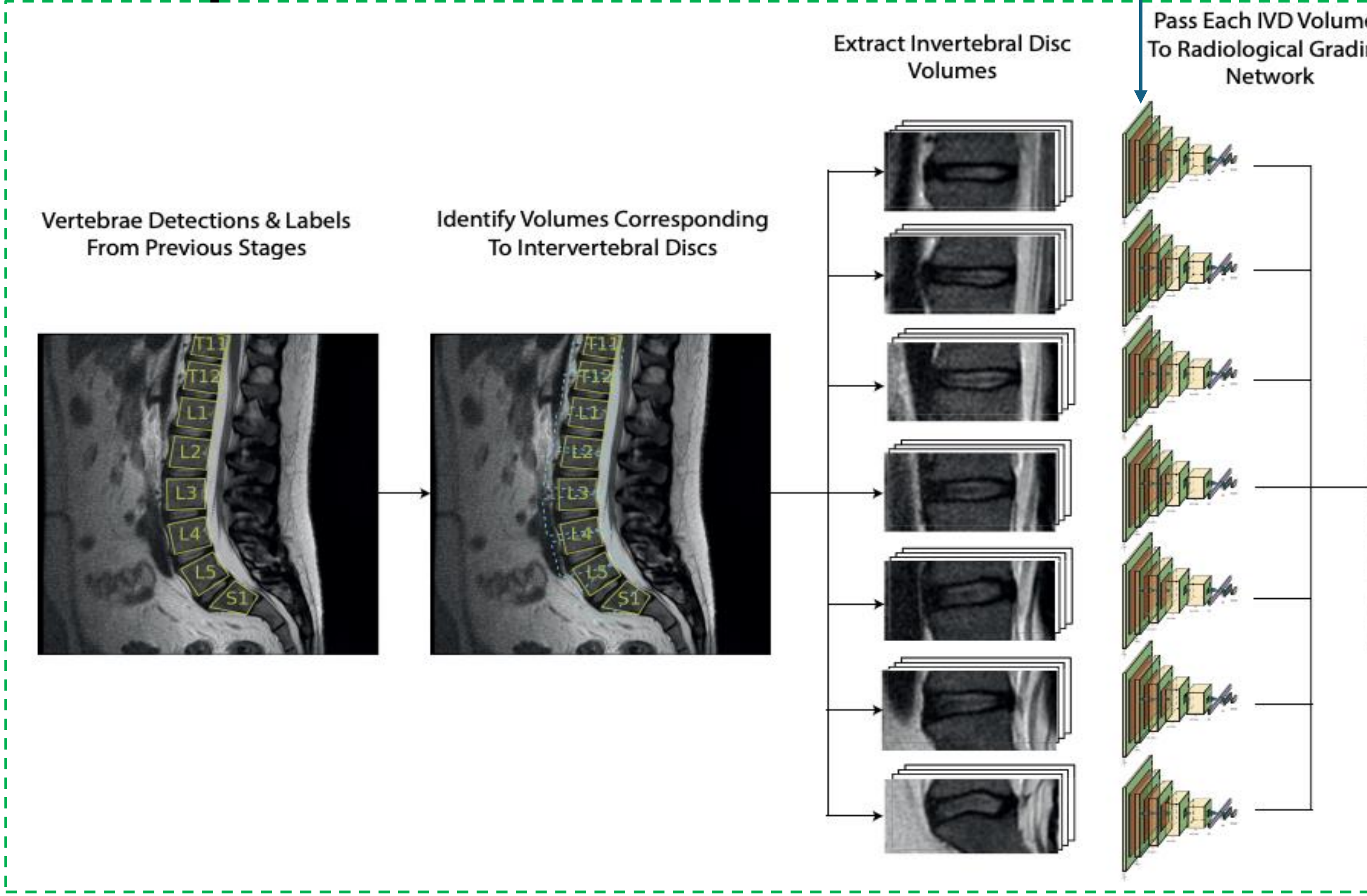
Radiological Features Evaluated

1. Disc degeneration (Pfirrmann grade)
2. Disc Narrowing (height loss)
3. Central canal stenosis
4. Spondylolisthesis
5. Endplate defects (upper and lower)
6. Marrow/Modic changes (upper and lower)
7. Foraminal stenosis (left and right)
8. Herniation



SpineNetV2

Conventional
ResNet32 model



	Pfirrmann	Narrowing	CentralCanalStenosis	Spondylolisthesis	UpperEndplateDefect	LowerEndplateDefect
T11-T12	1	1	1	0	0	0
T12-L1	1	1	1	0	0	0
L1-L2	1	1	1	0	0	0
L2-L3	2	1	1	0	0	0
L3-L4	2	1	1	0	0	0
L4-L5	2	1	1	0	0	0
L5-S1	1	1	1	0	0	0

	UpperMarrow	LowerMarrow	ForaminalStenosisLeft	ForaminalStenosisRight	Herniation
T11-T12	0	0	0	0	0
T12-L1	0	0	0	0	0
L1-L2	0	0	0	0	0
L2-L3	0	0	0	0	0
L3-L4	0	0	0	0	0
L4-L5	0	0	0	0	0
L5-S1	0	0	0	0	0

Dataset , Cohort and & Validation Pipeline

- **Sample size:** 1,747 lumbosacral discs collected from 353 patients (with a mean age of 54 $\pm$ 15.4 years, and 44.5% female).
- **Imaging used:** Sagittal T2-weighted MRI
- **Ground truth:** Consensus grading by 2 expert radiologists
- **Exclusion:** Incomplete scans, artefacts

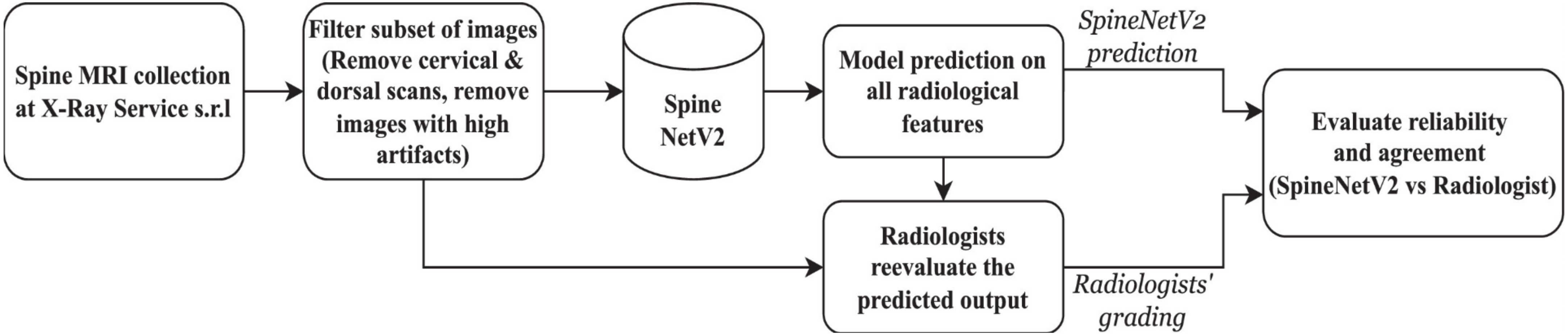
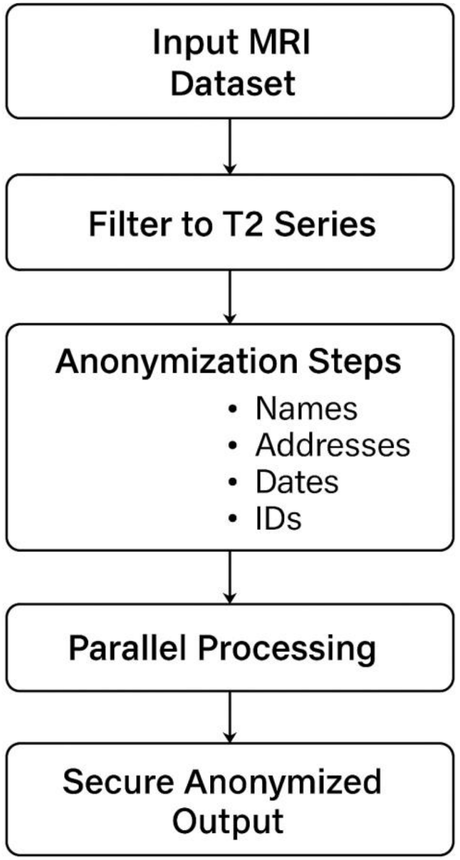


Fig: Summary of steps in the external validation process used in this study.

De-identification of patients data

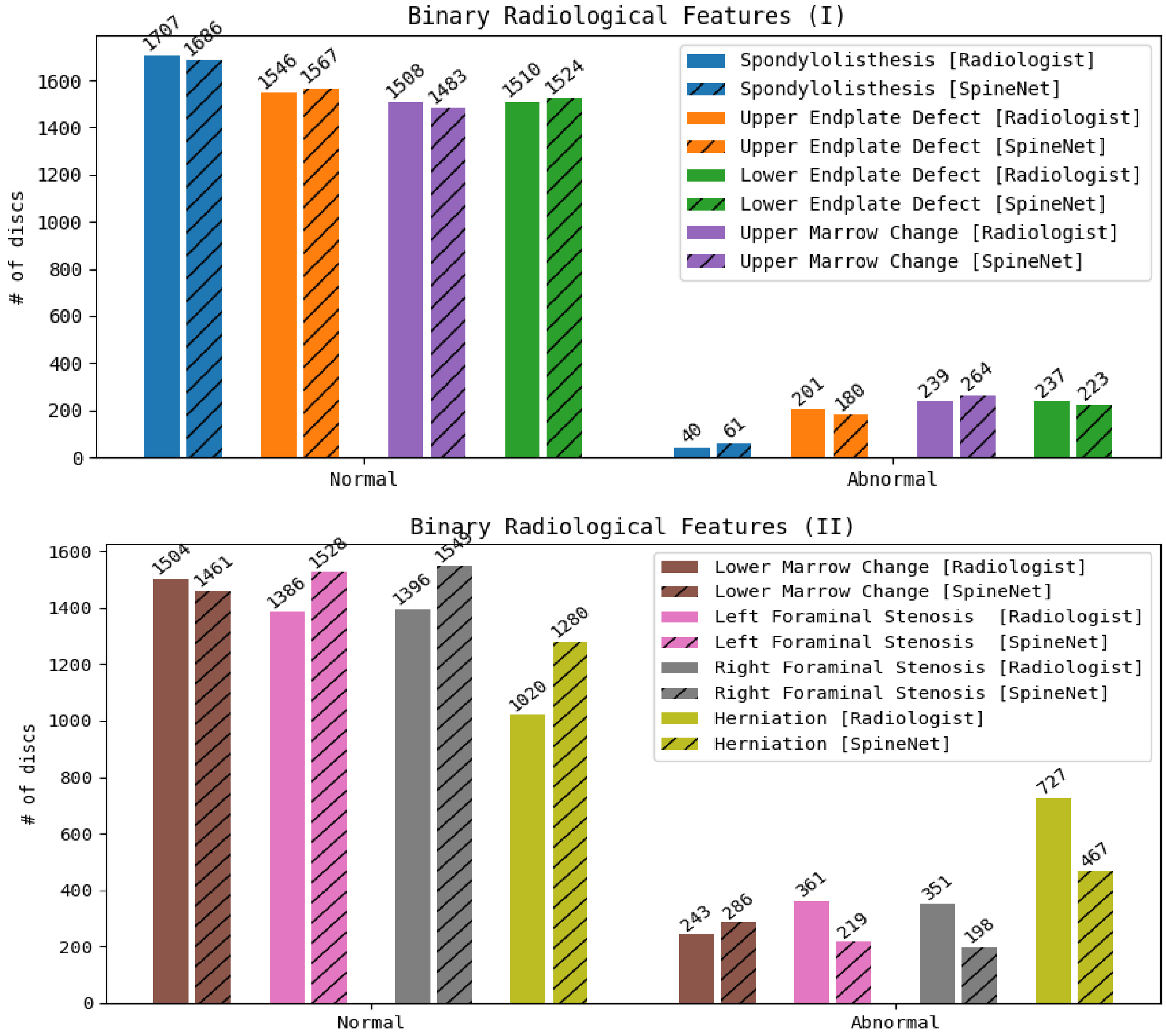
GDPR & HIPAA-Compliant DICOM Anonymization

- **Metadata Scrubbing:**
  Removes names, addresses, dates, private tags & unique identifiers.
- **Data Mapping & Transformation:**
  •*PatientID, PatientSex, PatientAge* → Rewritten with randomized or padded values.
  •Sequential attributes like RequestAttributesSequence are deleted.
- **Selective Targeting:**
  Processes only *T2_AX* and *T2_SAG* folders for relevant clinical imaging.
- **Parallel Processing:**
  Uses multiprocessing to handle 900+ patient folders efficiently.
- **Checksum Validation:**
  Confirms DICOM structural integrity post-anonymization.
- **Logging & Audit Trail:**
  Logs all actions to *anonymization.log* for traceability and error tracking.



Code available on: https://github.com/AlexSisay/Dicom_deidentify.git

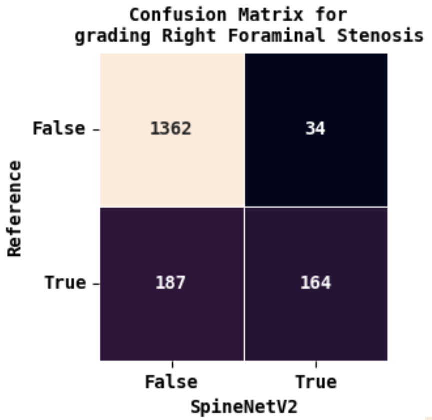
Overall Performance of the validation process

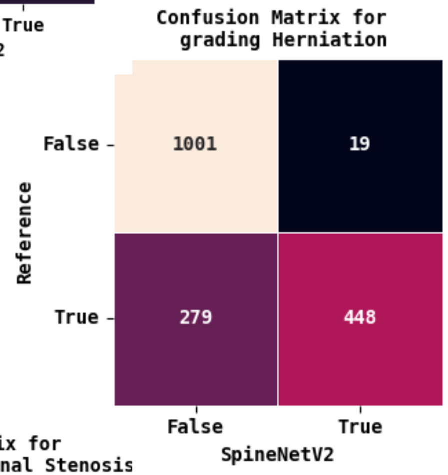


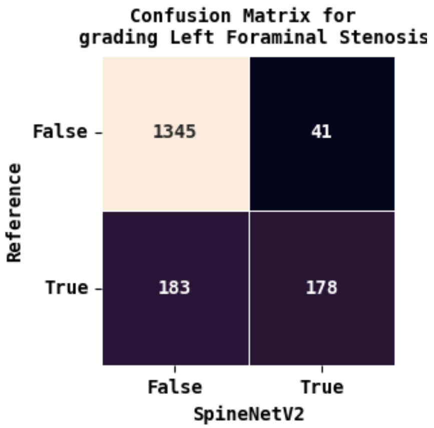
Feature-wise grading agreement

Metrics	Pfirrmann	Narrowing	Central Canal Stenosis	Spondylolisthesis	Upper endplate Defect	Lower Endplate Defect	Upper Marrow Change	Lower Marrow Change	Right Foraminal Stenosis	Left Foraminal Stenosis	Herniation
Accuracy	0.796	0.867	0.971	0.983	0.948	0.942	0.940	0.931	**0.852**	**0.854**	**0.790**
BAS	0.794	0.849	0.779	0.988	0.849	0.863	0.895	0.892	**0.702**	**0.691**	**0.759**
Precision	0.799	0.869	0.971	0.983	0.946	0.940	0.943	0.937	**0.841**	**0.843**	**0.810**
Recall	0.796	0.868	0.971	0.983	0.948	0.942	0.940	0.931	**0.852**	**0.854**	**0.790**
F1 Score	0.796	0.868	0.971	0.985	0.947	0.941	0.941	0.933	**0.838**	**0.838**	**0.779**
BSL	0.081	0.066	0.014	0.016	0.052	0.058	0.060	0.070	**0.148**	**0.146**	**0.210**
κ	0.738	0.799	0.749	0.705	0.732	0.745	0.756	0.731	**0.473**	**0.457**	**0.546**
LCCC	0.952	0.972	0.932	0.745	0.880	0.920	0.896	0.873	**0.542**	**0.530**	**0.630**
MCC	0.738	0.799	0.749	0.721	0.733	0.745	0.757	0.734	**0.494**	**0.483**	**0.576**

Where: BAS - Balanced Accuracy Score, BSL - Brier score loss, κ - Cohen's kappa, LCCC - Lin's Concordance Correlation Coefficient, MCC- Mathews correlation coefficient.







Conclusion & Future Directions

	Pfirrmann	Narrowing	Central Canal Stenosis	Spondylolisthesis	Upper Endplate Defect	Lower Endplate Defect	Upper Marrow	Lower Marrow	Foraminal Stenosis Left	Foraminal Stenosis Right	Herniation
SpineNet V2	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗

- **Strengths:** Robust generalizability across institutions, efficient processing (*~ 20 seconds per scan on CPU and only 3 seconds on GPU)*.
- **Limitations:** Variability in **foraminal stenosis** and **herniation** grading; need for integration of axial scans.
- **Clinical Relevance:** Reduces reporting time, supports large-scale research.
- **Next Steps:** Integration of axial scans, integration with radiological workflows.

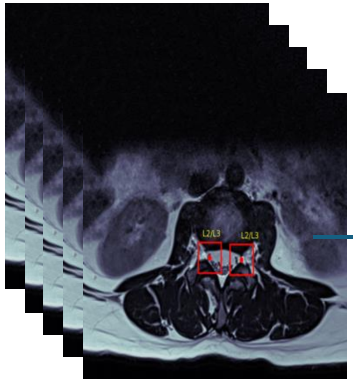
Nigru et. al. *External validation of SpineNetV2 on a comprehensive set of radiological features for grading lumbosacral disc pathologies, North American Spine Society Journal (NASSJ), 2024, DOI: https://doi.org/10.1016/j.xnsj.2024.100564*

So, what further refinement should be made?

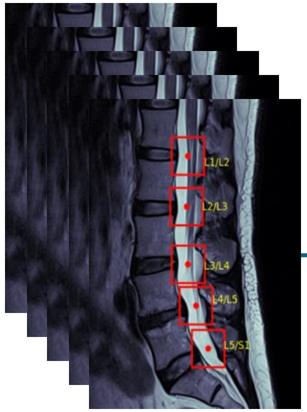
- Integrate axial view MR into the AI pipeline

On Progress!

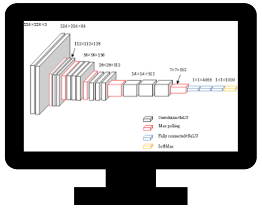
Axial T1/T2



Sagittal T1/T2



AI Models



Radiological grade

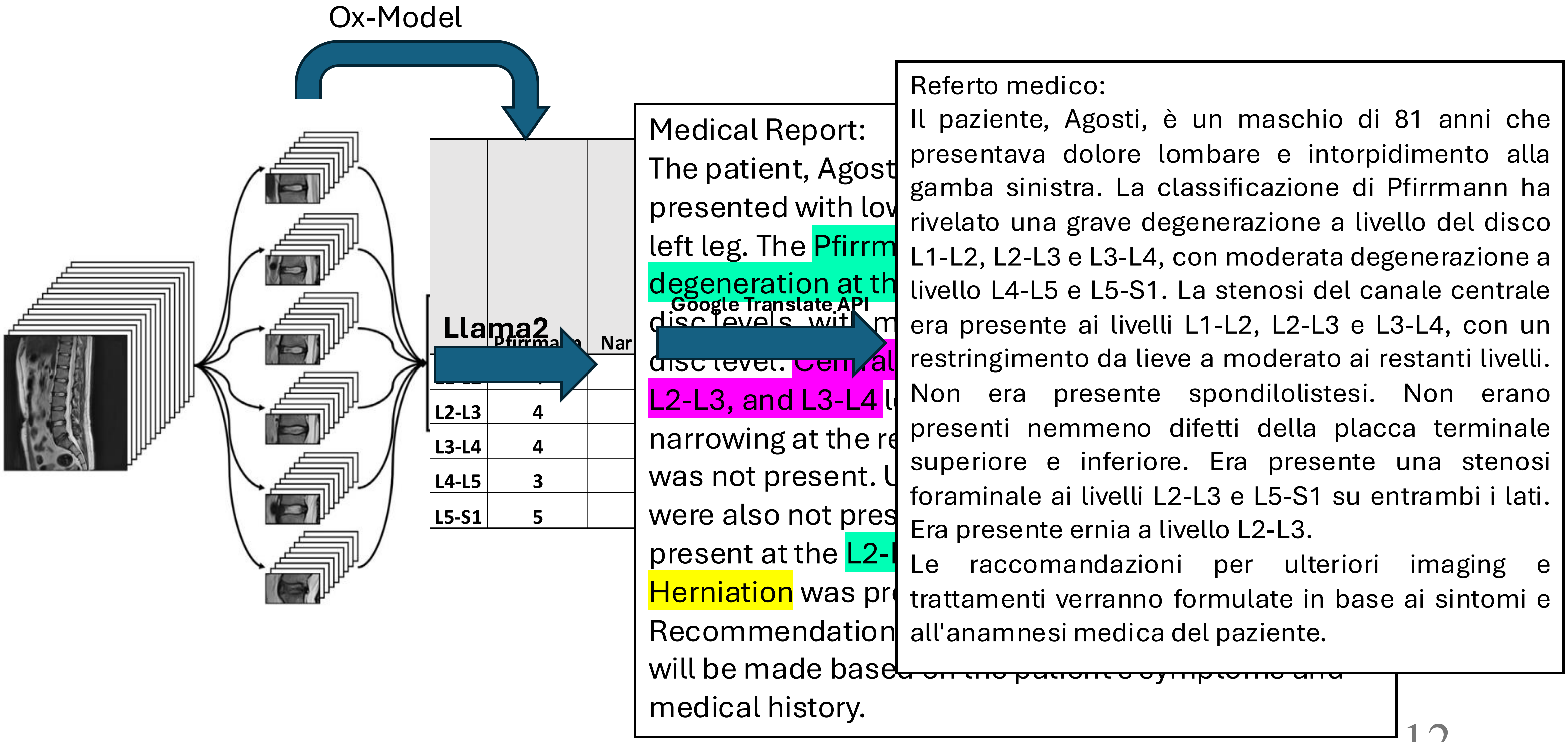
LLM

Radiologist Report

- AI based Lumbar spine discs classification using Sagittal + axial view T1/T2w MRI

http://127.0.0.1:7860

SOTA Review + end to end use case



X-Ray Service

UNIVERSITÀ DEGLI STUDI DI BRESCIA

2. AI-Driven Exploration of Patient-Reported Outcomes in Chronic Spinal Pain:

From EDA to Machine Learning with Psychosocial and Disability Metrics

Motivation and Background

- Chronic back pain: a multidimensional health challenge.
- Psychological, social, and disability components often underrepresented in traditional analyses.
- Role of AI: uncover hidden patterns, patient clusters, and actionable insights.

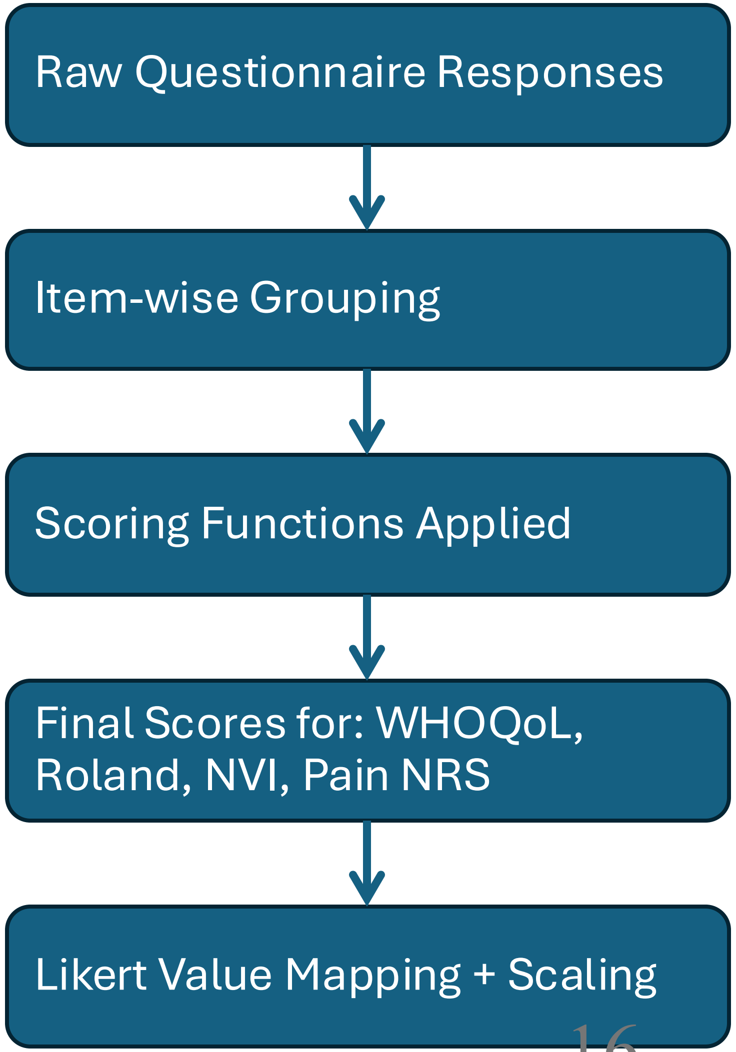
Dataset Overview

- Data has been collected from 113 consenting patients, with collection ongoing.
- **4 primary instruments**:
  - Roland-Morris Disability Questionnaire (23 items)
  - WHOQoL-BREF (Physical, Psychological, Social, Environmental)
  - COPE NVI (Coping Orientation to Problems Experienced Nuova versione Italiana) -- Coping Strategy
  - Pain NRS

Raw Questionnaire to Scoring Process

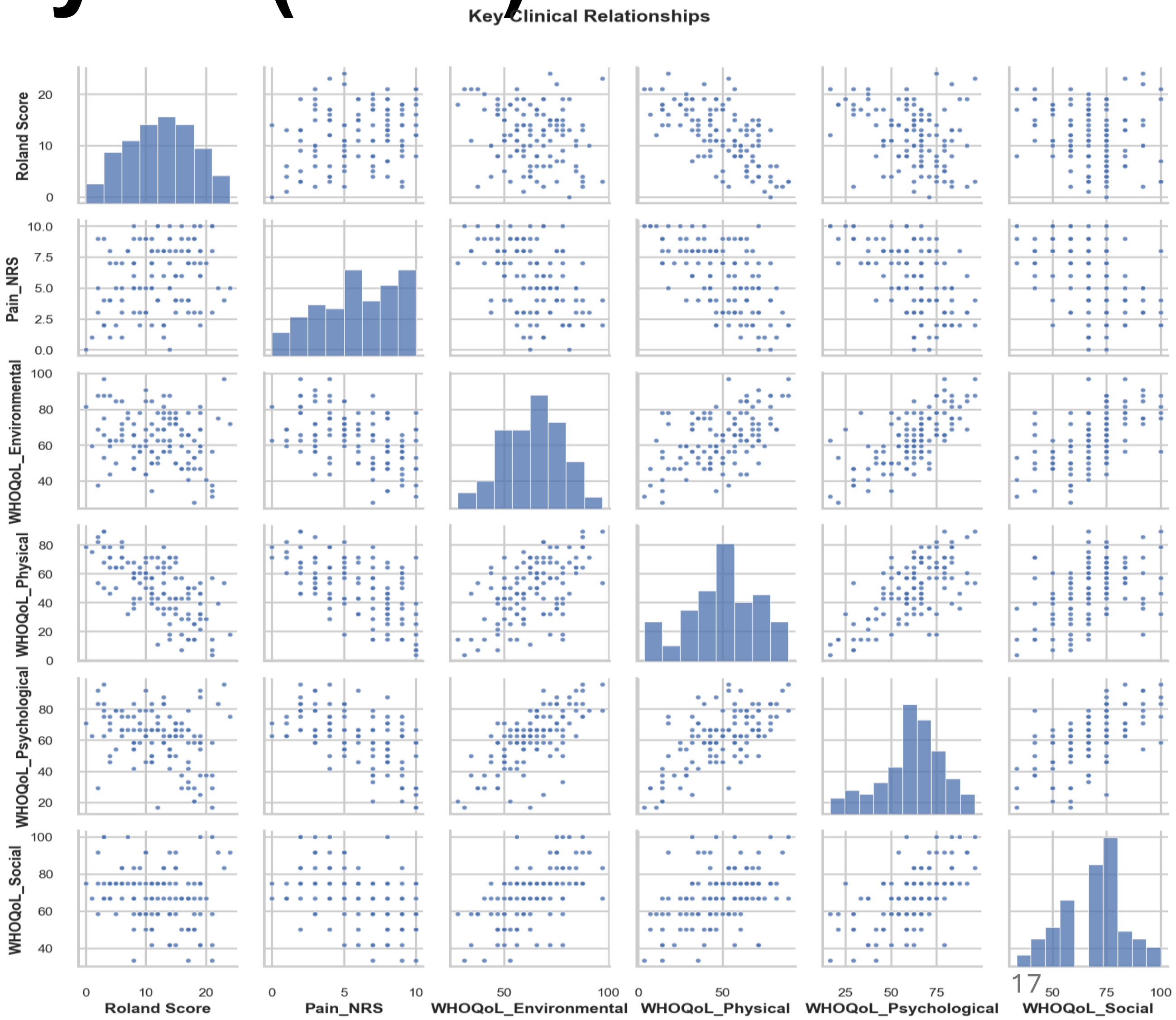
Each domain score is computed from **item-level responses** using official scoring procedures:

Instrument	Items	Score Range	Notes
Roland-Morris	24	0–24	Sum of items; higher = greater disability
WHOQoL-BREF	26	0–100 per domain	Transformed to 0–100 scale per WHO manual; Higher = Better QoL
Coping Strategies (NVI-25)	25	1–5 Likert per strategy	Mean of grouped items per strategy; Higher = Mostly used strategy
Pain NRS	1	1-10	1-10; higher = severe pain



Exploratory Data Analysis (EDA)

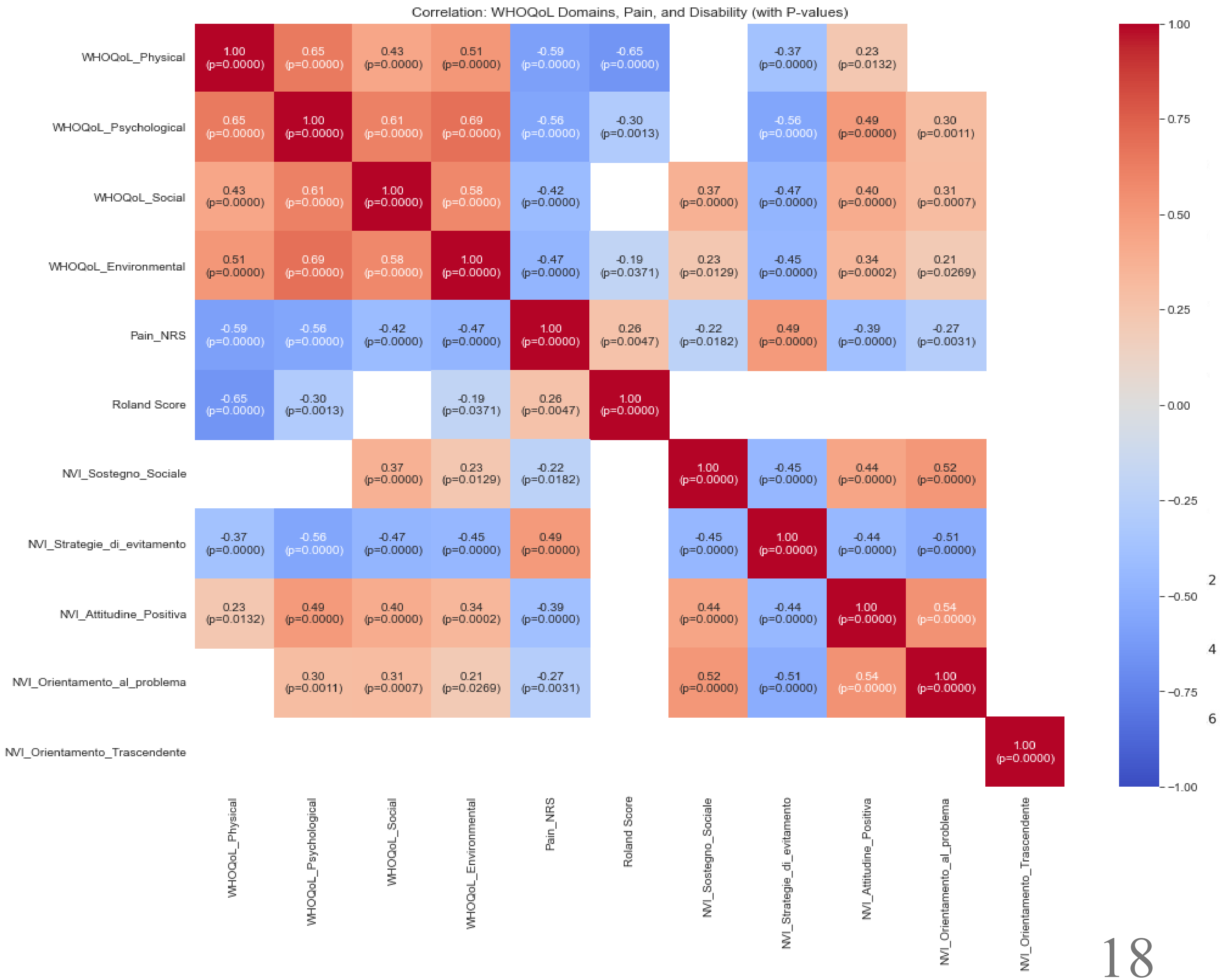
- **Distributions** of Roland Score, WHOQoL dimensions, Pain NRS.
- Pairwise **correlations**



X-Ray Service

Correlation Heatmap

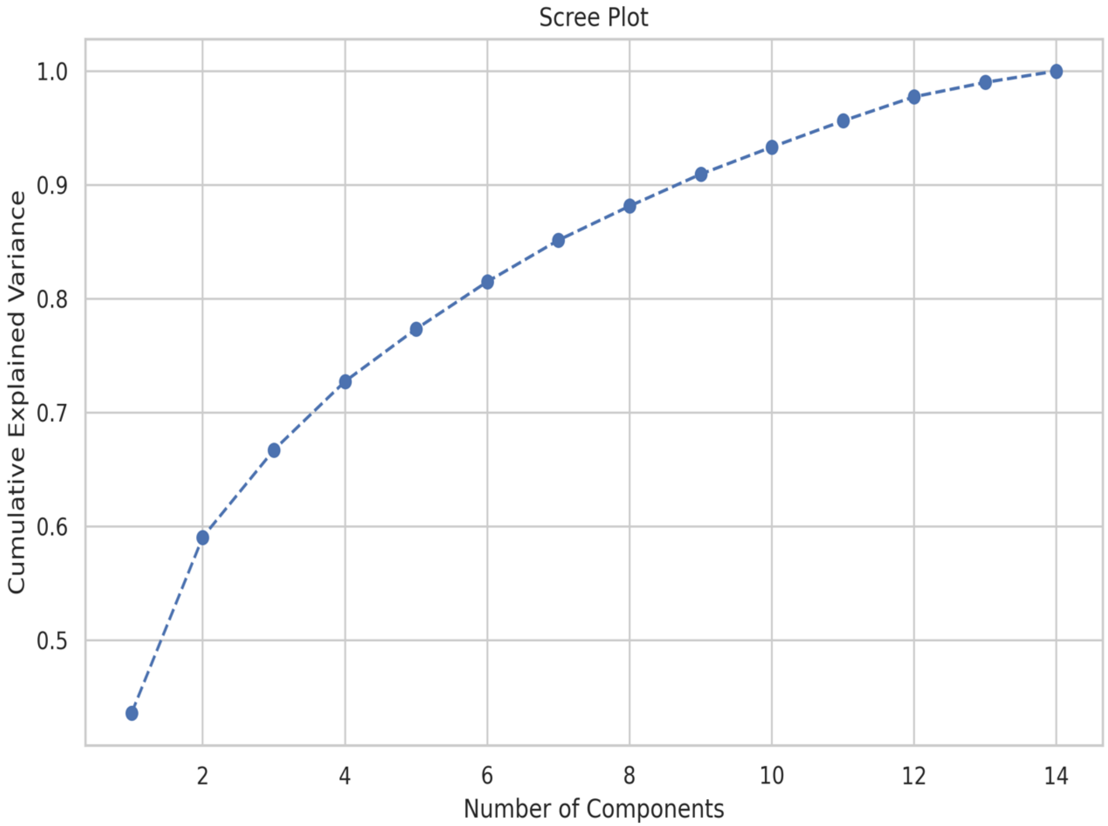
UNIVERSITÀ DEGLI STUDI DI BRESCIA

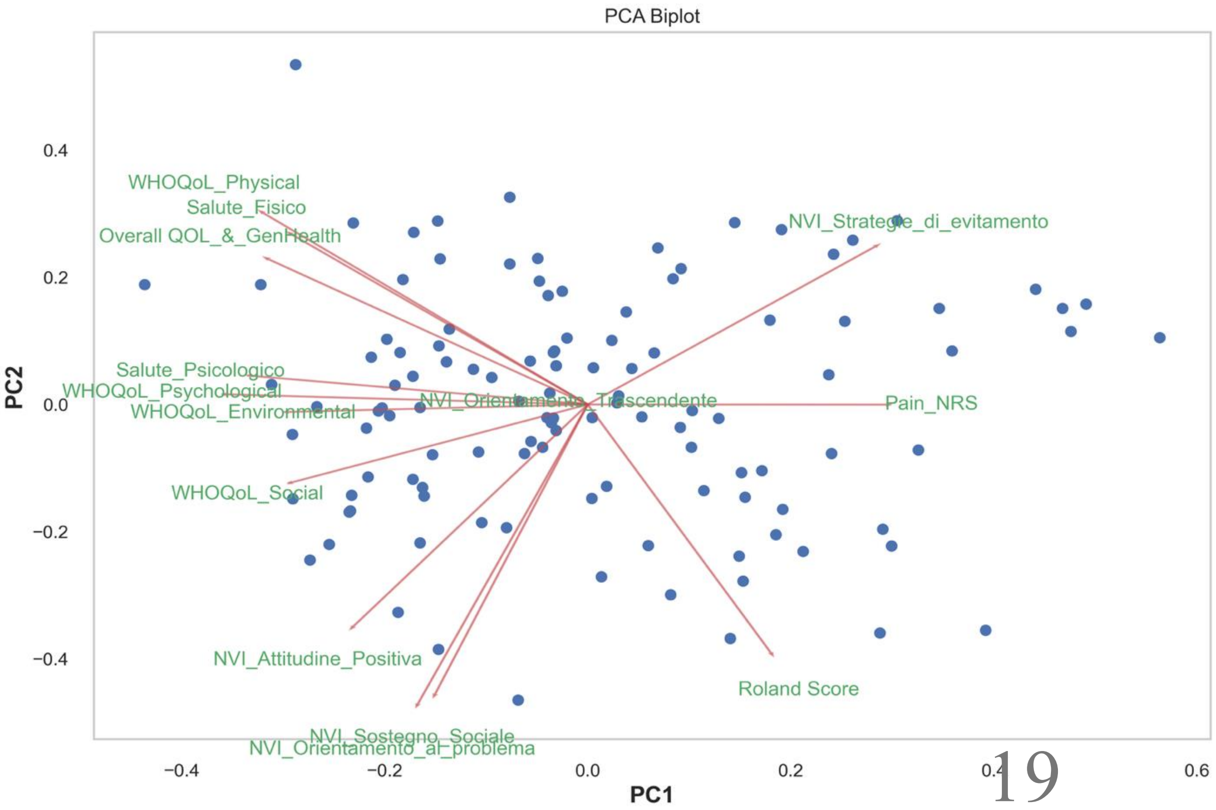


Significance mask applied:
p_values >= 0.05

Dimensionality Reduction – PCA

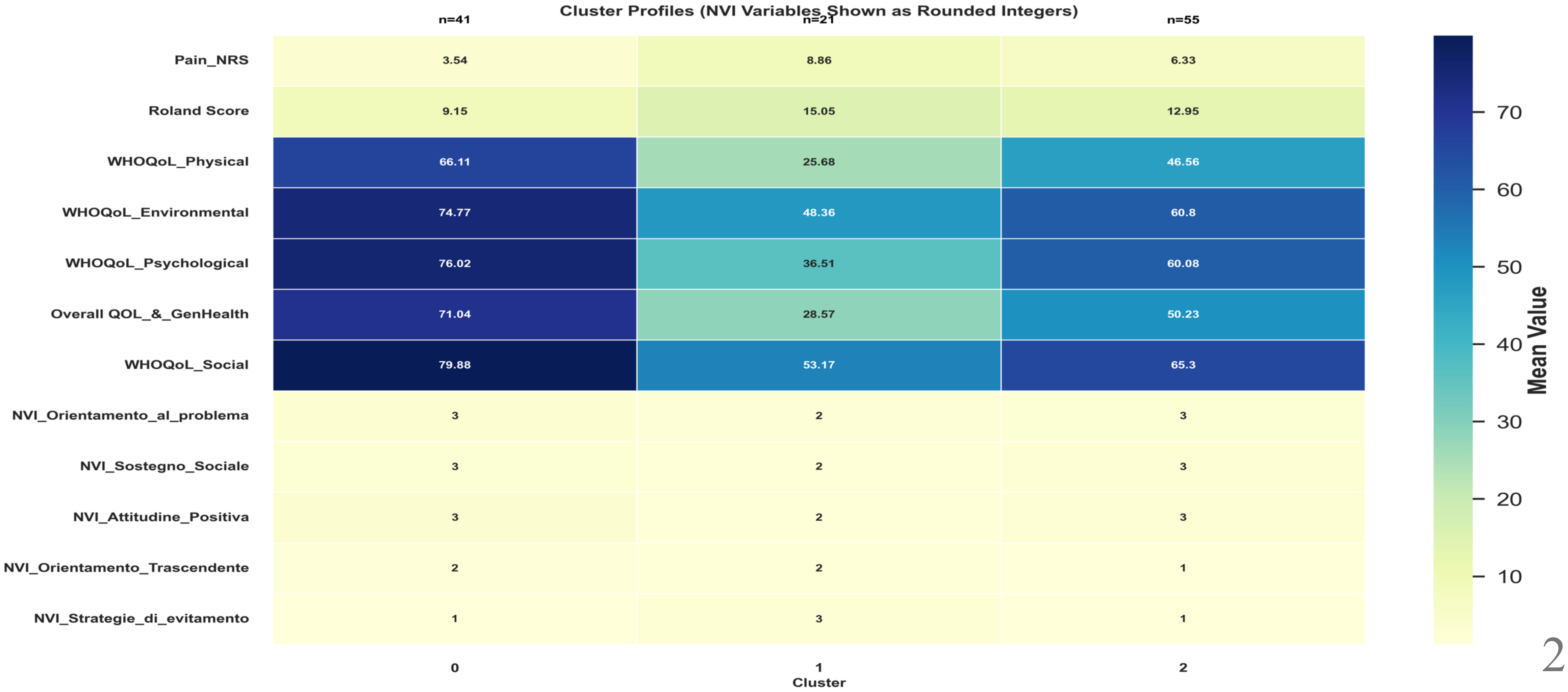
- Dimensionality Reduction – PCA
  - PCA to reduce noise and explore latent structure
  - Top 2 PCs explain variance across Roland, WHOQoL, and NVI variables





Unsupervised Learning (Clustering)

- KMeans clustering on PCA features (k=3 clusters)



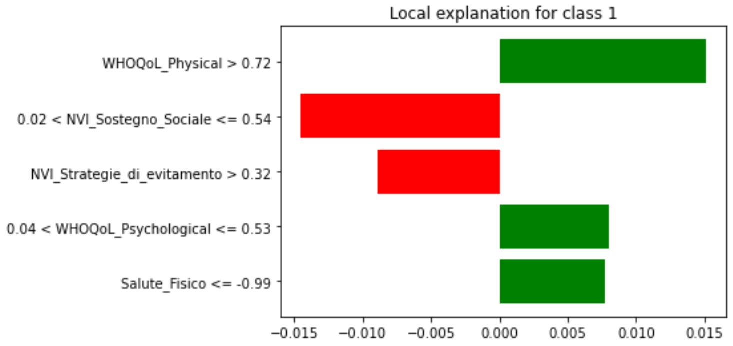
X-Ray Service

UNIVERSITÀ DEGLI STUDI DI BRESCIA

Works on progress:

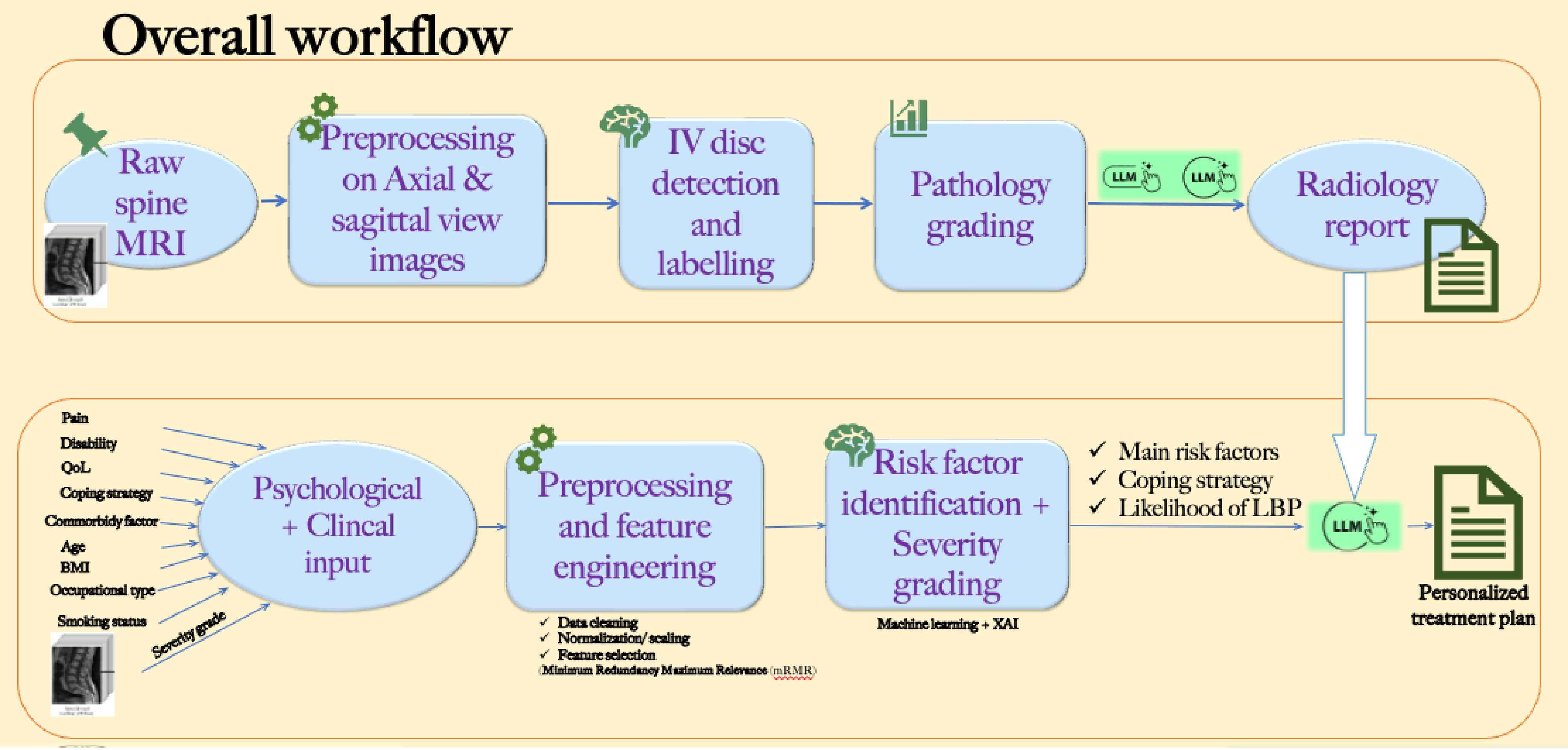
- Add severity grades as a new variable and train a supervised ML model to predict the severity level (from their psycho-social + clinical data).
- Add a feature of explainability to identify the relevant features of the model prediction (xAI)
- Integration with the previous image model and introduce personalized treatment plan.

LIME example



- This helps in **debugging** the model’s behavior and help us identifying main risk factors.
- We can easily spot if the model is relying on **irrelevant features**.
- Supports **model transparency** for regulatory and trust requirements (like **GDPR**, **HIPAA**).

End to End architecture



Thank You!