Prediction of End-Organ Complications of Type 2 Diabetes: A Machine Learning Approach

Luca Romeo*, Michele Bernardini°, Emanuele Frontoni*

°Università Politecnica delle Marche

*University of Macerata

luca.romeo@unimc.it, m.bernardini@univpm.it, emanuele.frontoni@unimc.it

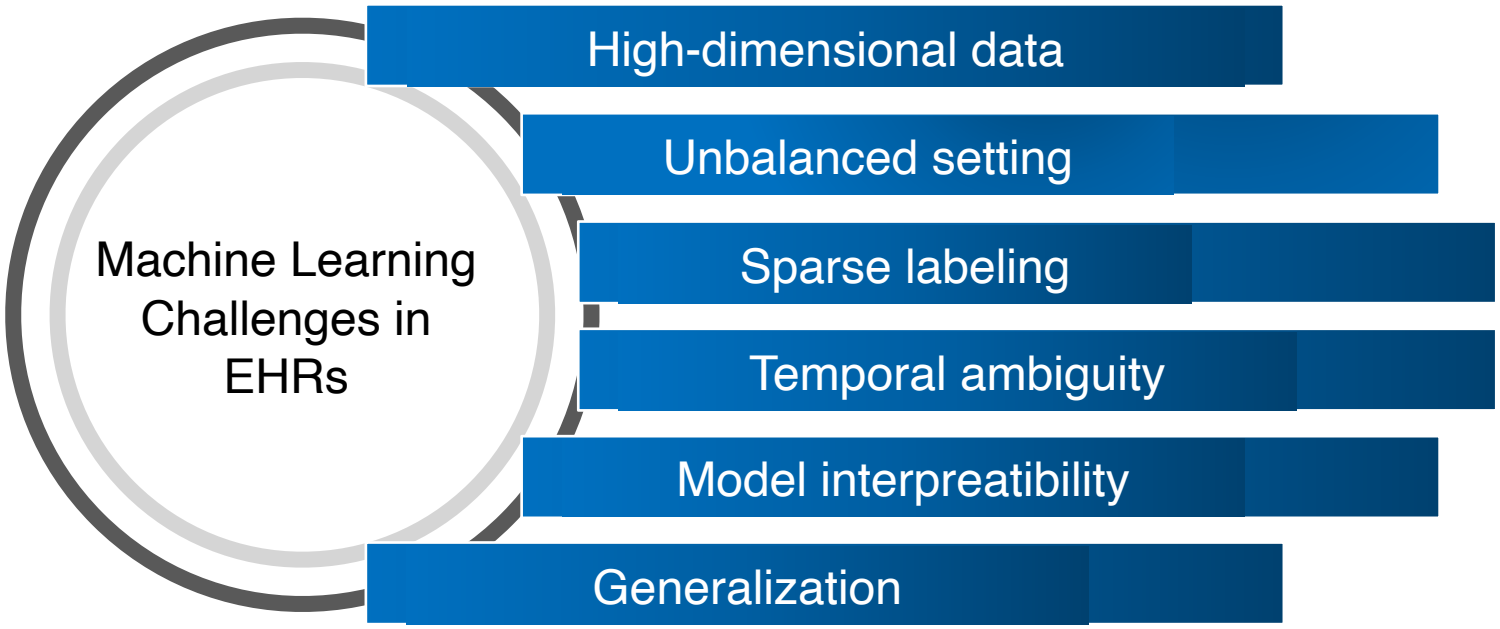
SWETALY

unimc l'umanesimo che innova

VRAI vision robotics artificial intelligence

UNIVERSITÀ POLITECNICA DELLE MARCHE

Challenges



Romeo, L., Armentano, G., Nicolucci, A., Vespasiani, M., Vespasiani, G., & Frontoni, E. (2020, July). A Novel Spatio-Temporal Multi-Task Approach for the Prediction of Diabetes-Related Complication: a Cardiopathy Case of Study. In ***IJCAI***(pp. 4299-4305).

Romeo, L., & Frontoni, E. (2022). A Unified Hierarchical XGBoost model for classifying priorities for COVID-19 vaccination campaign. ***Pattern Recognition***, *121*, 108197.

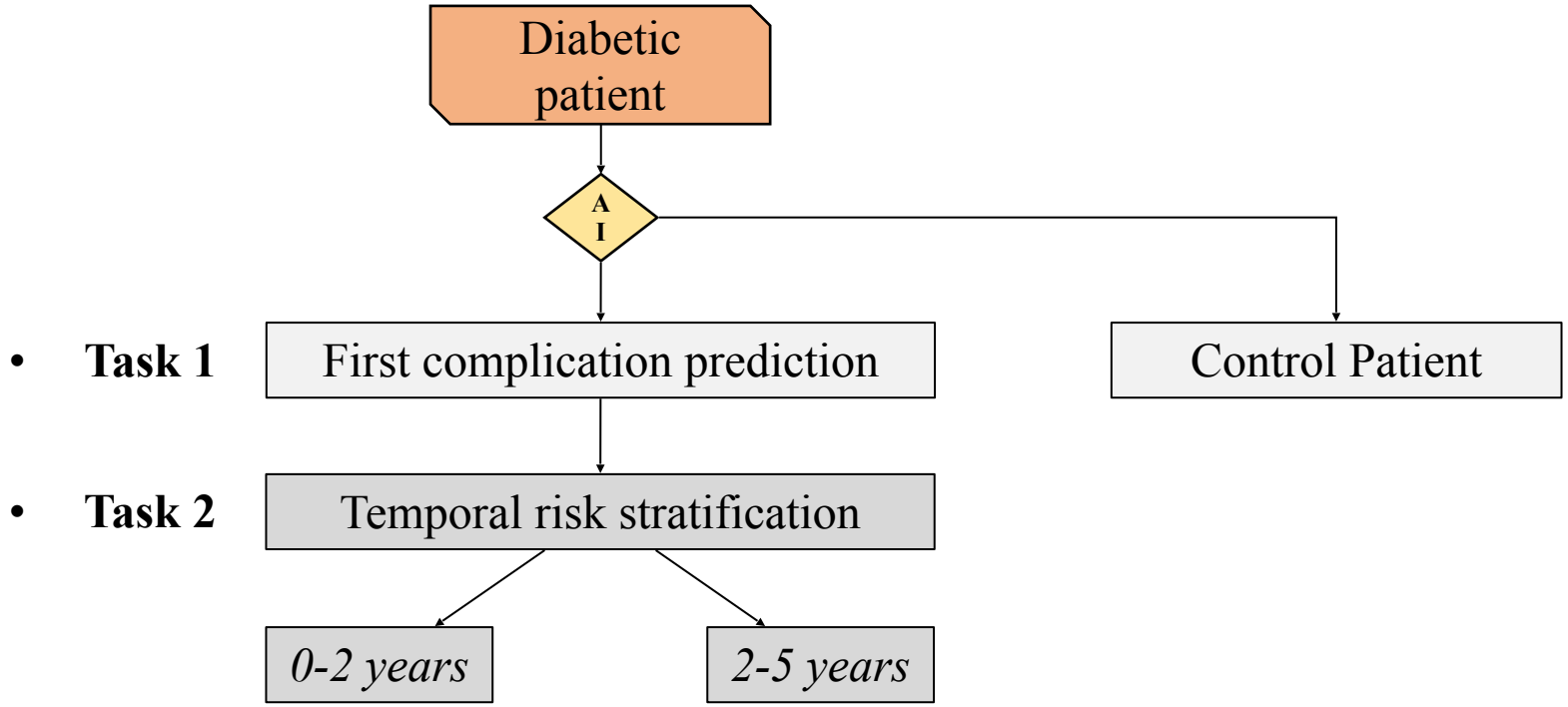
Bernardini, M., Romeo, L., Frontoni, E., & Amini, M. R. (2021). A semi-supervised multi-task learning approach for predicting short-term kidney disease evolution. ***IEEE Journal of Biomedical and Health Informatics***, *25*(10), 3983-3994.

Bernardini, M., Morettini, M., Romeo, L., Frontoni, E., & Burattini, L. (2020). Early temporal prediction of type 2 diabetes risk condition from a general practitioner electronic health record: a multiple instance boosting approach. ***Artificial Intelligence in Medicine***, *105*, 101847.

Bernardini, M., Romeo, L., Misericordia, P., & Frontoni, E. (2019). Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine. ***IEEE Journal of Biomedical and Health Informatics***, *24*(1), 235-246.

Objective

First complication risk prediction in diabetic patients and temporal risk stratification

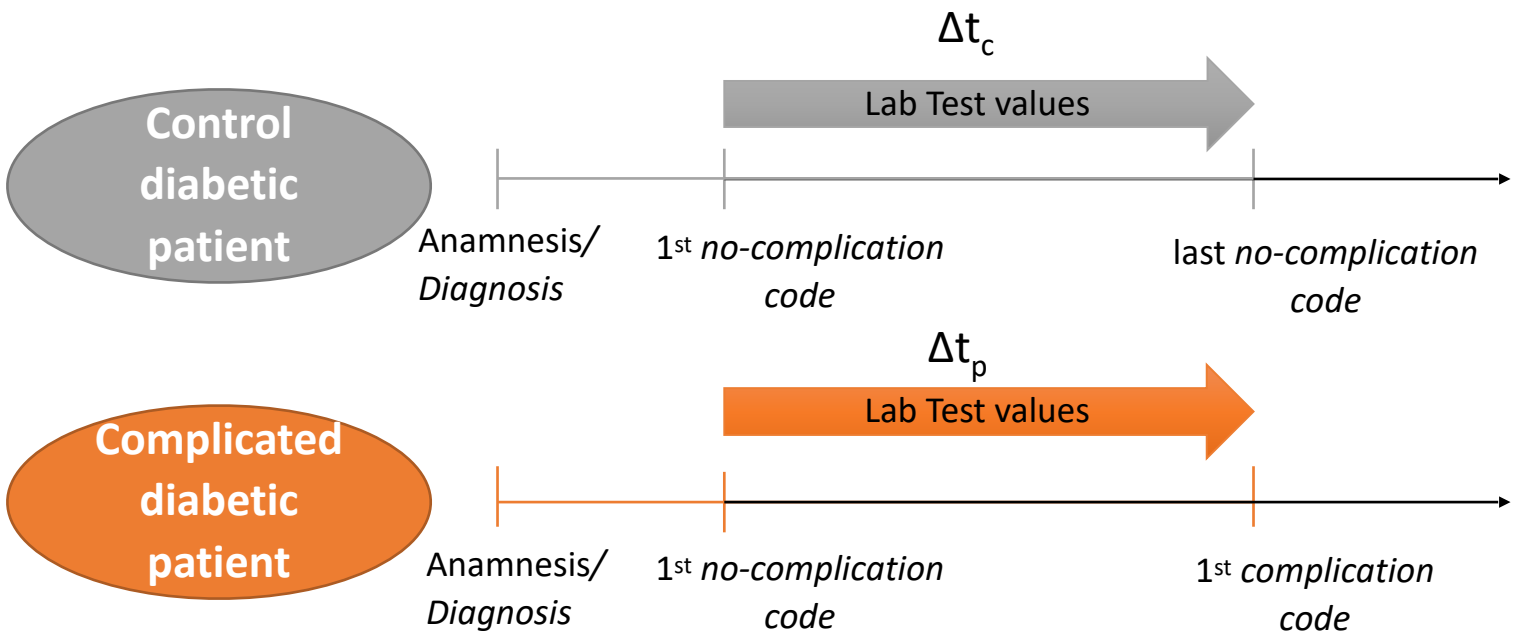


Complications:

- Retinopathy
- Cer. Vasculopathy
- Per. Vasculopathty
- Cardiopathy
- Nefropathy
- Neuropathy

*M. Bernardini, L. Romeo, A. Mancini and E. Frontoni, "A Clinical Decision Support System to Stratify the Temporal Risk of Diabetic Retinopathy," in **IEEE Access**, vol. 9, pp. 151864-151872, 2021*

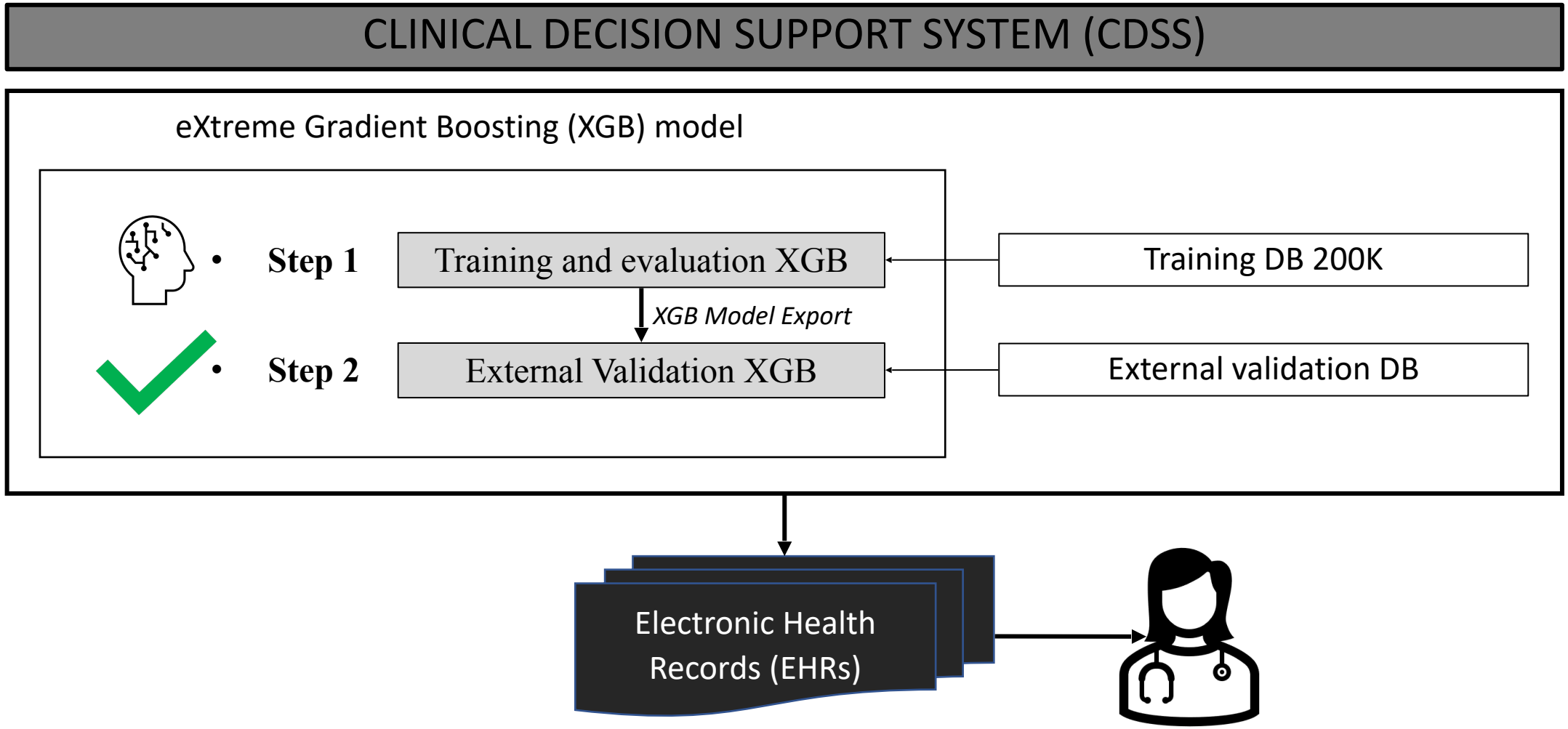
Preprocessing: observational time-window



Only prescribed exams in the range of the selected observational time-window were considered:

- Prescribed exams between the first no-complication code and the last no-complication code for **Control** diabetic patients
- Prescribed exams between the first no-complication code and the first complication code for **Complicated** diabetic patients

XGB model: Training/Evaluation and External Validation



Training and evaluation XGB: Statistics

vision robotics artificial intelligence

Step 1: Training DB 200K

147.664 Diabetic patients

↓

Preprocessing: Inclusion/exclusion criteria

↓

DB training 200K	Total patients	Control pt.	Compl. pt.
Retinopathy	40555	31611	8944
Cer. Vasculopathy	7852	4883	2969
Per. Vasculopathy	9314	7769	1545
Cardiopathy	17849	12305	5544
Nephropathy	30510	22593	7917
Neuropathy	9827	7751	2076

	Lab tests description	Uom		Lab tests description	Uom
1	Glicemia pre-prandiale	mg/dl	33	Trigliceridi post 12h dig.	mg/dl
2	Glicemia post-prandiale	mg/dl	34	Sodio (Uri)	mEq/L
3	ACR	mg/mmol	35	Albuminuria/creatinuria	NULL
4	Clearance Creatinina	ml/min	36	Proteine (Uri)	mg/dl
5	Colesterolo LDL (calc)	mg/dl	37	Piastrine	1000/mm3
6	Creatininuria	mg/dl	38	Microalbuminuria	mg/l
7	Emoglob.Glicata HBA1C (Lab.3)	%	39	Glicosuria	G/LITRO
8	Emoglob.Glicata HBA1C (Lab.2)	%	40	Glicemia a digiuno	mg/dl
9	Emoglob.Glicata HBA1C (Lab.1)	%	41	Gamma GT	UI/L
10	Indice di Winsor sx	NULL	42	Fosfatasi Alcalina	UI/L
11	Indice di Winsor dx	NULL	43	Fibrinogeno (San)	mg/dl
12	ACR (Calc)	mg/mmol	44	Emoglobina A1 (totale)	NULL
13	Microalbuminuria	mg/24h	45	Emoglobina	g/dl
14	BMI	Kg/m²	46	Emoglob.Glicata HbA1c	%
15	Urino coltura 1=neg 2=pos	NULL	47	Creatinina Clearance Uri F	ml/min
16	Potassio urinario	mEq/l	48	Creatinina Clearance Uri M	ml/min
17	AER III	mcg/min	49	Creatinina	mg/dl
18	AER II	mcg/min	50	Creatin Fosfo Chinasi (Sie)	UI/L
19	Microalbuminuria (II)	mg/l	51	Colesterolo LDL	mg/dl
20	Glicemia prima di pranzo	mg/dl	52	Colesterolo HDL	mg/dl
21	Glicemia prima di cena	mg/dl	53	Colesterolo	mg/dl
22	Glicemia ore 23	mg/dl	54	Circonferenza vita	cm
23	Glicemia dopo pranzo	mg/dl	55	SGOT	UI/L
24	Glicemia dopo colazione	mg/dl	56	Amilasi (Uri)	U/L
25	Glicemia dopo cena l	mg/dl	57	Amilasi	UI/L
26	Clearance Creatinina (calc)	ml/min	58	AER	mcg/min
27	Chetoni Urine	mg/dl	59	GPT	UI/L
28	Pressione Diastolica	mmHg	60	Ac. Urico	mg/dl
29	Pressione Sistolica	mmHg	61	Sesso	NULL
30	Altezza	cm	62	Età	anni
31	Peso	kg	63	Durata diabate	anni
32	Urea	mg/dl	64	Sequenza Id soggetto	NULL

Training and evaluation XGB: Challenges & Outputs

vision robotics artificial intelligence

Step 1: Training DB 200K

- **Challenges:**
  - Redundant predictors → Intrinsic features selection in the XGB model
  - High-rate of missing values in predictors → Data imputation
  - Control/Complicated diabetic patients imbalance → SMOTE (Oversampling)
  - False Negative patients minimization → Adaptative thresholds of prediction probabilities

- **Outputs:**
  - Prediction probabilities
  - Features importance (XGB model interpretability/explainability)
  - XGB model export

Training and evaluation XGB: Results TenfoldCV – Task1

vision robotics artificial intelligence

Step 1: Training DB 200K

- **Task 1** First complication prediction

AI model	Area Under the curve (AUC)	Sensitivity (True Positive)	False Positive	Specificity (True Negative)	False Negative
Retinopathy					
XGBoost	85.81	**82.15**	17.85	**72.38**	27.62
Cer. Vasculopathy					
XGBoost	85.55	**82.26**	17.74	**71.75**	28.25
Per. Vasculopathy					
XGBoost	84.19	**83.54**	16.46	**67.56**	32.44
Cardiopathy					
XGBoost	87.29	**71.06**	28.94	**85.47**	14.53
Nephropathy					
XGBoost	91.95	**81.83**	18.17	**86.66**	13.34
Neuropathy					
XGBoost	84.32	**83.14**	16.86	**68.37**	31.63

Negative: Control diabetic patient

Positive: Complicated diabetic patient

Training and evaluation XGB: Results TenfoldCV – Task2

vision robotics artificial intelligence

Step 1: Training DB 200K

- **Task 2** — Temporal risk stratification

AI model	Area Under the Curve (AUC)	Sensitivity (2-5 years)	False Positive	Specificity (0-2 years)	False Negative
Retinopathy					
XGBoost	79.20	**63.73**	36.27	**77.00**	23.00
Cer. Vasculopathy					
XGBoost	81.35	**72.07**	27.93	**73.86**	26.14
Per. Vasculopathy					
XGBoost	80.22	**67.42**	32.58	**75.95**	24.05
Cardiopathy					
XGBoost	80.10	**65.00**	35.00	**77.97**	22.03
Nephropathy					
XGBoost	82.28	**68.68**	31.32	**79.75**	20.25
Neuropathy					
XGBoost	80.57	**71.43**	28.57	**72.37**	27.63

Long-term risk → Complication in 2-5 years

Mid-term risk → Complication in 0-2 years

External Validation XGB: Results - Task 1

vision robotics artificial intelligence

Step 2: External validation DB

- **Task 1** First complication prediction **(ID4 center)**

Control pt. / Complicated pt	AUC	Sensitivity (TP)	False Positive	Specificity (TN)	False negative
Retinopathy					
5374/2250	88.62	**81.16**	18.84	**81.65**	18.35
Cer. Vasculopathy					
1520/782	66.50	**63.55**	36.45	**58.31**	41.69
Per. Vasculopathy					
508/268	81.99	**73.13**	26.87	**76.57**	23.43
Cardiopathy					
4651/1623	81.18	**73.01**	26.99	**70.72**	29.28
Nephropathy					
1544/385	97.86	**90.91**	9.09	**98.32**	1.68
Neuropathy					
1297/587	85.80	**73.94**	26.06	**83.65**	16.35

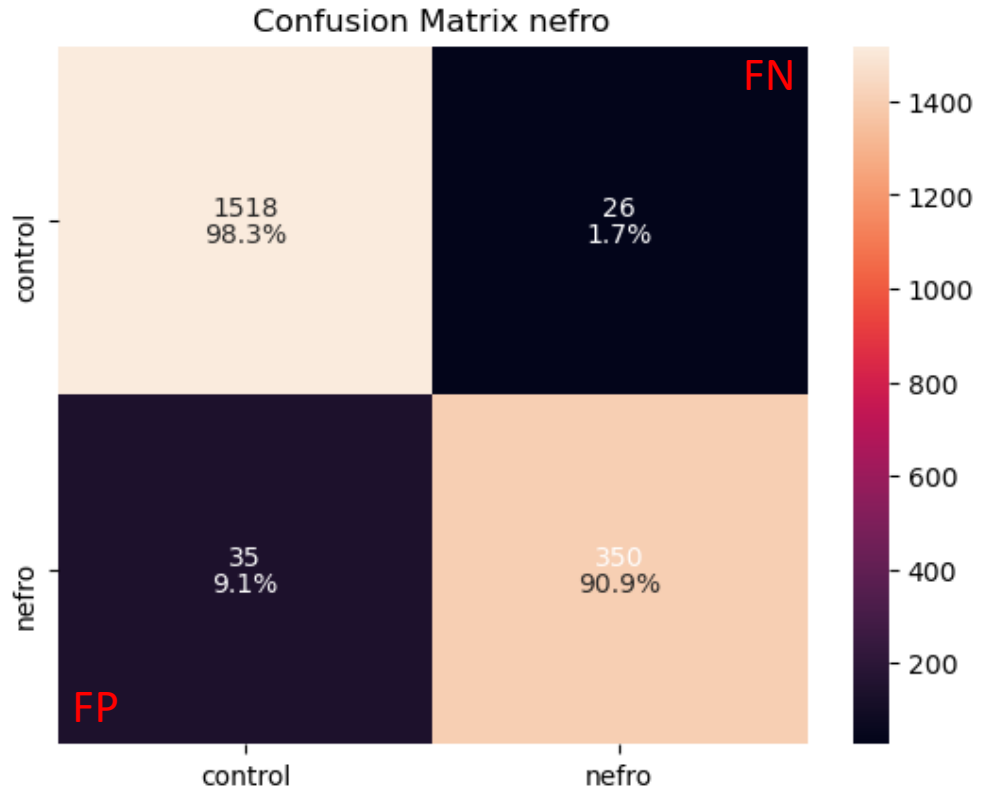
External Validation XGB: Results - Task 1 – Nephropathy

vision robotics artificial intelligence

Step 2: External validation DB

- **Task 1** First complication prediction – **Nephropathy Example (ID4 center)**

Ranking	Importance	Lab test (predictors)	Missing values
#1	4.54 %	Microalbuminuria	1.35 %
#2	4.43 %	Uric acid	52.98 %
#3	4.29 %	Creatinin	35.72 %
#4	4.05 %	LDL cholesterol	62.42 %
#5	3.83 %	Fasting glycaemia	2.18 %
#6	3.74 %	Gamma GT	56.40 %
#7	3.74 %	Glycated hemoglobin HbA1c	0.41 %
#8	3.71 %	Triglycerides	37.12 %
#9	3.59 %	Systolic pressure	52.26 %
#10	3.57 %	HDL Cholesterol	40.80 %



External Validation XGB: Results - Task 2

vision robotics artificial intelligence

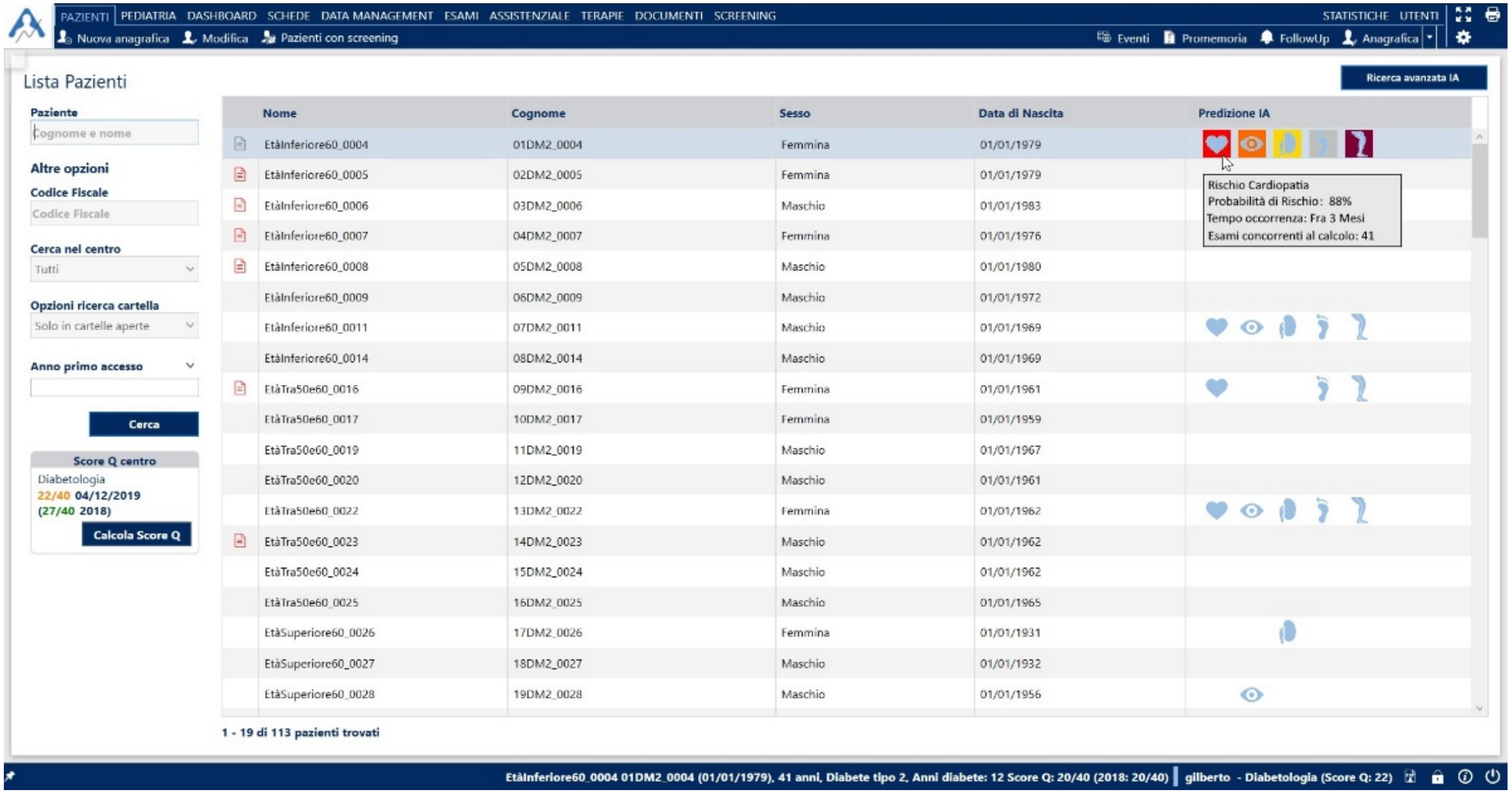
Step 2: External validation DB

- **Task 2** — Temporal risk stratification **(ID4 center)**

Model	AUC	Specificity 0-2 yy	False negative 0-2 yy	Sensitivity 2-5 yy	False positive 2-5 yy
Retinopathy					
XGBoost	95.29	**93.07**	6.93	**83.86**	16.14
Cer. Vasculopathy					
XGBoost	88.55	**91.94**	8.06	**63.40**	36.60
Per. Vasculopathy					
XGBoost	90.62	**92.06**	7.94	**67.24**	32.76
Cardiopathy					
XGBoost	91.37	**92.99**	7.01	**68.69**	31.31
Nephropathy					
XGBoost	93.94	**98.34**	1.66	**64.65**	35.35
Neuropathy					
XGBoost	92.83	**95.73**	4.27	**63.49**	36.51

CDSS Software integrated in EHR

vision robotics artificial intelligence



Thank you for the attention

Luca Romeo*, Michele Bernardini°, Emanuele Frontoni*

°Università Politecnica delle Marche

*University of Macerata

luca.romeo@unimc.it, m.bernardini@univpm.it, emanuele.frontoni@unimc.it

SWETALY

unimc l'umanesimo che innova

VRAI vision robotics artificial intelligence

UNIVERSITÀ POLITECNICA DELLE MARCHE