Consiglio Nazionale delle Ricerche

BUILD-IT2023 Workshop
October 19-20 2023, Rome

Towards a Digital Twin for Personalized Diabetes Prevention: The PRAESIIDIUM Project

Alessia Paglialonga, Marta Lenatti, Davide Simeone, Pierluigi Francesco De Paola, Alberto Carlevaro, Maurizio Mongelli, Fabrizio Dabbene
Cnr-Istituto** di **Elettronica** e di **Ingegneria** dell'**Informazione** e delle **Telecomunicazioni (CNR-IEIIT)

Filippo Castiglione, Maria Concetta Palumbo, Paola Stolfi, Alessandro Ravoni, Paolo Tieri
Cnr-Istituto** per le **Applicazioni** del **Calcolo (CNR-IAC)

Contacts: alessia.paglialonga@cnr.it; paolo.tieri@cnr.it

Co-funded by the European Union

This work was supported in part by the European Union through the project PRAESIIDIUM ''Physics Informed Machine Learning-Based Prediction and Reversion of Impaired Fasting Glucose Management'' (call HORIZON-HLTH-2022-STAYHLTH-02) under Grant 101095672

PRAESIIDIUM PROJECT

Co-funded by the European Union

PRAESIIDIUM - PHYSICS INFORMED MACHINE LEARNING-BASED PREDICTION AND REVERSION OF IMPAIRED FASTING GLUCOSE MANAGEMENT

Project Idea

To develop a tool aimed at providing a **real-time prediction of the prediabetic risk** of an individual.

The prediction algorithm will be based on a "*physics-informed machine learning*" approach: a rich dataset of real-life data, obtained from already existing previous and a new clinical trial with continuous data ingestion through wearable sensors, will be **combined with mathematical models and eXplainable AI (XAI) techniques**, to overcome the limits of "*black-box*" ML approaches, while improving the prediction performances and reducing the computational time of the risk calculation based on the simulations of the mathematical model.

The final algorithm will be implemented in **a web-based platform**, where medical doctors and patients can inject data from several sources (acquisition form connected sensors and manual insertion) and obtain a **real time analysis of the risk** to develop the prediabetic condition over time.

PRAESIIDIUM will develop easy-to-use, reliable, efficient and innovative tools (e.g., web/ mobile app, mathematical models, wearable sensors) for the **personalized management and real-time prediction of the prediabetic risk**. This goal will result in benefits for affected citizens, their families, hospitals, different industries, scientific community, governments/ policy-makers, and environment.

PRAESIIDIUM

Duration
01/01/2023 – 31/12/2025

Project Number
101095672

11 Participants from 7 countries

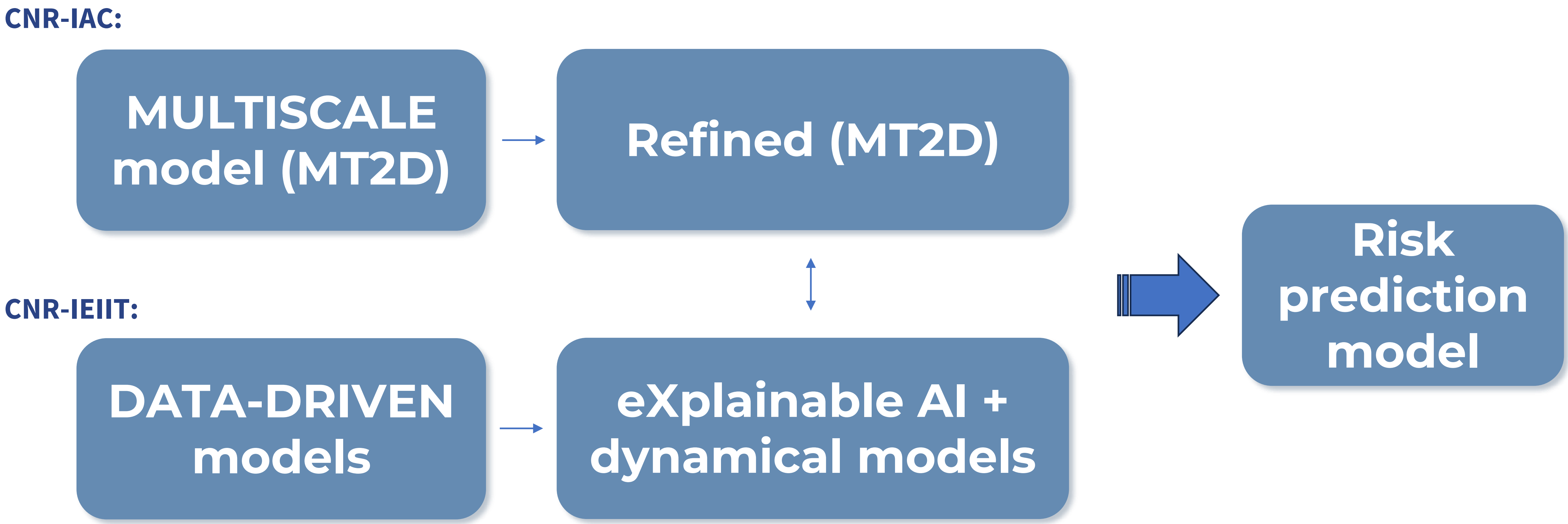
https://praesiidium.spindoxlabs.com/

@praesiidium

praesiidium

Consiglio Nazionale delle Ricerche

PERSONALIZED RISK PREDICTION AND PREVENTION



MULTISCALE MODEL MT2D

Whole-body, multi-scale model of metabolic homeostasis during exercise

An agent-based simulator of the human immune system

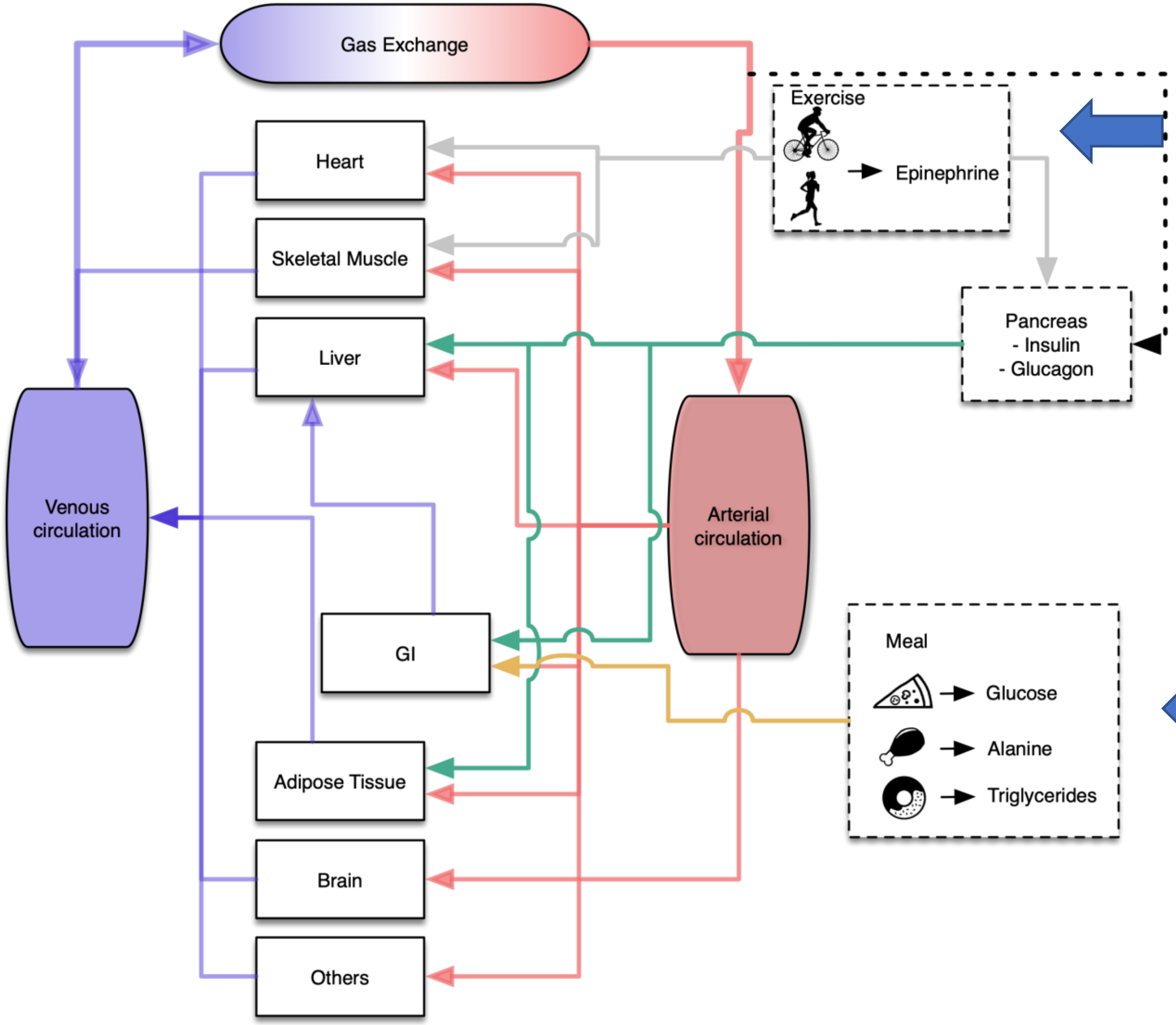
F. Castiglione, F. Celada. Immune system modelling and simulations. Boca Raton: CRC Press; 2015

J. Kim, G. M. Saidel, and M. E. Cabrera. Multi-scale computational model of fuel homeostasis during exercise: effect of hormonal control. *Annals of Biomedical Engineering*, 2007, 35: 69-90.

M.C. Palumbo, M. Morettini, P. Tieri, F. Diele, M. Sacchetti, and F. Castiglione. Personalizing physical exercise in a computational model of fuel homeostasis. *PLoS Computational Biology,* 2018, 14: e1006073.

M.C. Palumbo, A.A. de Graaf, M. Morettini, P. Tieri, S. Krishnan, and F. Castiglione (2023). A computational model of the effects of macronutrients absorption and physical exercise on hormonal regulation and metabolic homeostasis. *Computers in Biology and Medicine*, 107158.

MULTISCALE MODEL MT2D – THE METABOLIC MODEL



Model of exercise

Model of glucagon/insulin

Model of food intake

142 ODEs

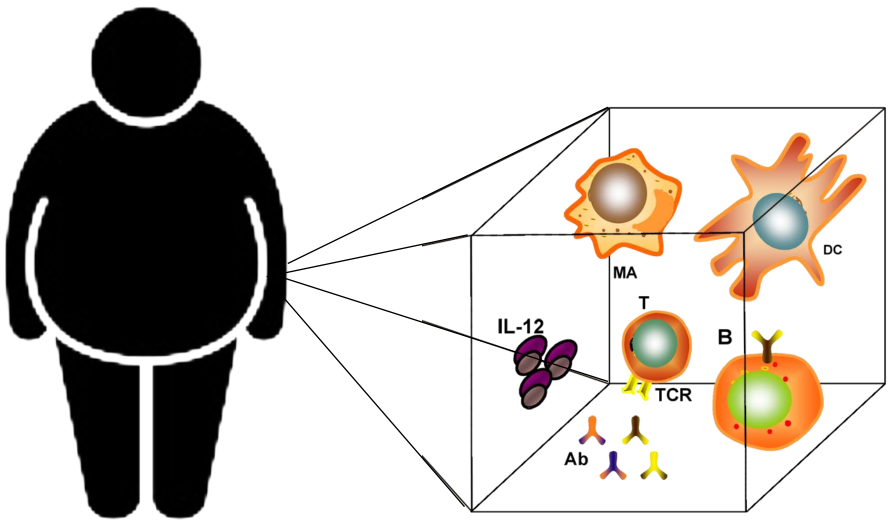
MULTISCALE MODEL MT2D – THE METABOLIC MODEL

Personalization of model predictions

Input Parameter	Meaning
S	Sex (male/female)
A	Age in years
BM	Body mass in kg
H	Height in m
VO_{2max}	Maximal oxygen uptake in $ml \cdot kg^{-1} \cdot min^{-1}$
T_v	Target value of exercise intensity in percentage of VO_{2max}
Fitness status	Cardio-respiratory fitness classification from “poor” to “superior”
$t_{ex,start}$, $t_{ex,end}$	Start/end of the exercise session in minutes
$C_{fast,glu}$	Fasting glucose level in mmol/l

MULTISCALE MODEL MT2D – THE IMMUNE SYSTEM

C-IMMSIM: an agent-based model of the immune system



Immune cells are individually represented. They follow behavioral rules coding for known immunological mechanisms

B (B-1, B-2), PLB, TH (Th1, Th2, Th17, Treg), CTL, Treg, NK, MA, DC

EP, ADIP

IgM, IgG (IgG1, IgG2), IC

IL-2, Danger, IL-12, IFN-g, IL-4, TNF-a, TGF-b, IL-10, IL-6, IL-10, IL-18, IL-23, IFN-b, IL-1b

LPS, Leptine

....

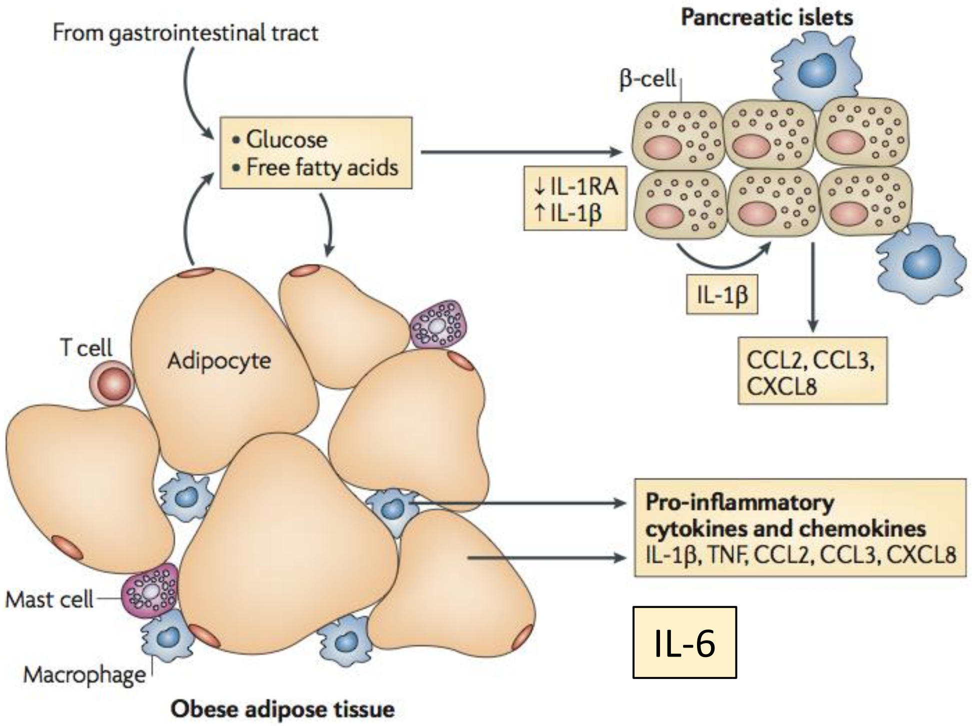
MULTISCALE MODEL MT2D – THE IMMUNE SYSTEM

Immunology of adipose tissue inflammation

- Adipose Tissue (AT) largely composed of adipocytes + pre-adipocytes, endothelial cells and immune cells

- In obesity AT undergoes **expansion** and **remodeling** (↑pro-, ↓anti-inflammatory cytokines and infiltration of immune cells)

- This obesity-associated pattern of events in AT → **chronic state of low-grade inflammation**



DATA-DRIVEN MODELS: RETROSPECTIVE DATASET

Canadian Primary Care Sentinel Surveillance Network (CPCSSN) EMR database (http://cpcssn.ca/)

Individual characteristics

- Age
- Sex
- BMI
- Height
- Weight
- …

Biomarkers

- Systolic blood pressure
- Diastolic blood pressure
- Fasting blood sugar
- A1C*
- Low density lipoprotein
- High density lipoprotein
- Total cholesterol
- Triglycerides
- Other biomarkers (tbd)

Comorbidities

- Hypertension
- Depression
- Chronic obstructive pulmonary disease
- Osteoarthritis
- Other conditions (to be defined)

Medications

- 1st line antidiabetics
- 2nd line antidiabetics
- Anti-hypertensive
- Anti-cholesterol
- Corticosteroids
- Antidepressant
- Smoking meds
- … other meds (tbd)

Risk factors

- Smoking
- Scarcely available: Alcohol; Physical activity; family history

Observation window: up to 1 year before possible onset

N = **64922** patients with T2D
N = **47271** patients with stable prediabetes
N = **23056** patients with one or more transitions (NG/PD/T2D)

N = **3885** patients with T2D with at least 1 full record available in the observation window (features: FBG, TG, TC, HDL, LDL, sBP, dBP, BMI, age, sex, no missing values). Of these, **2858** include information about smoking

DATA-DRIVEN MODELS: COUNTERFACTUAL EXPLANATIONS

GOAL: find minimum yet meaningful changes in biomarkers that may reduce the risk of developing T2D <u>on an individual basis</u>

DEFINITION:
Given an observation **x**= $(u_1,u_2,\ldots,u_n,z_1,z_2,\ldots,z_m)$, **x** $\in$ *class C*

Controllable features — Non-controllable features (e.g., age)

We want to determine the minimum variation of the controllable features Δu* so that:

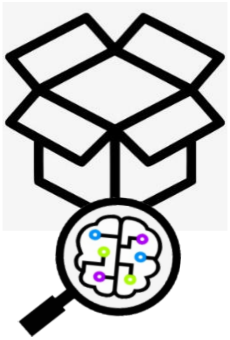
x*=(u| Δu*, z) **x*** $\in$ *class* C^*, with $C^* \neq C$

Input features

Age, Gender, Hypertension, sBP, BMI, LDL, HDL, TG, FBS

Output

- High risk T2D
- Low risk T2D



Increased transparency

Counterfactual Explanations

Application-grounded evaluation

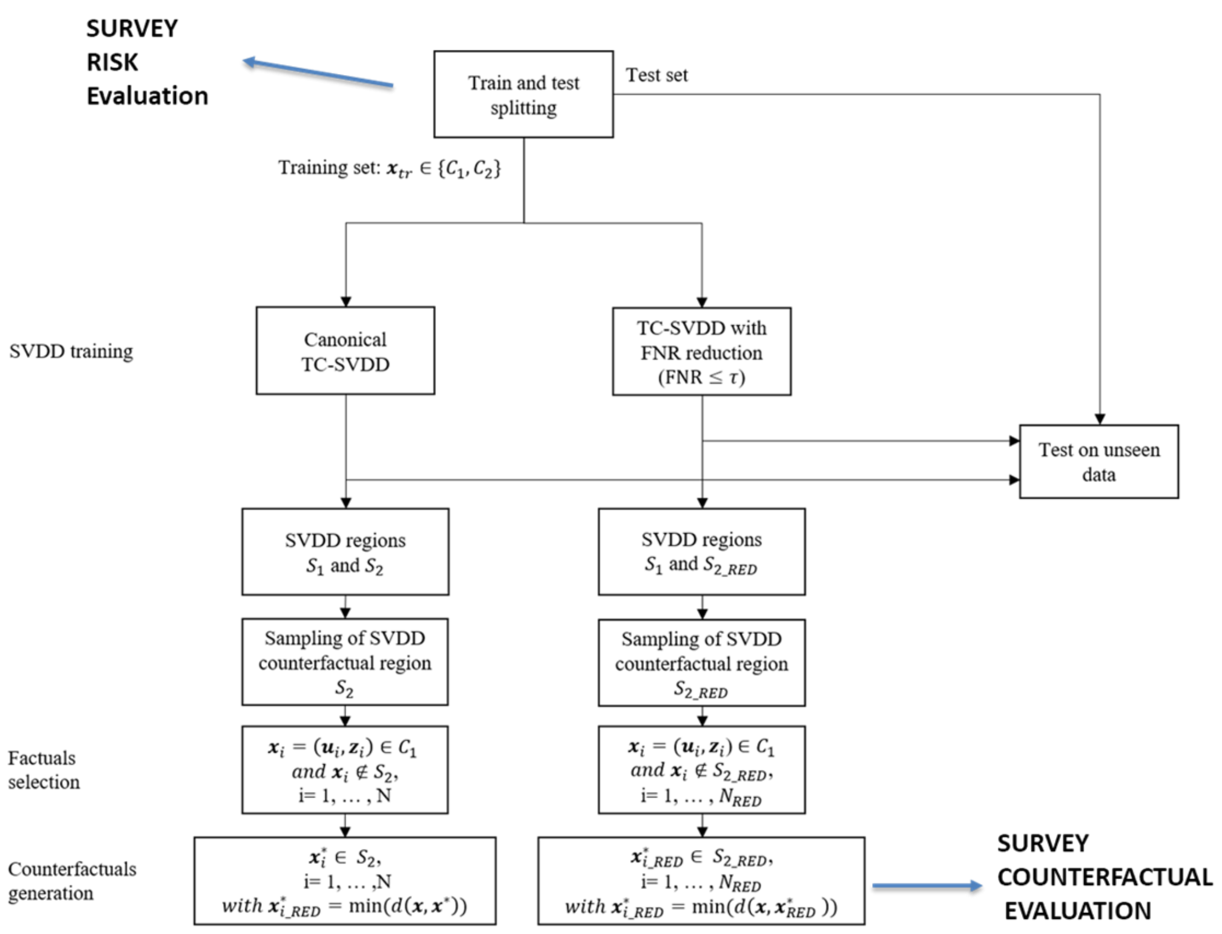
High risk T2D → low risk T2D

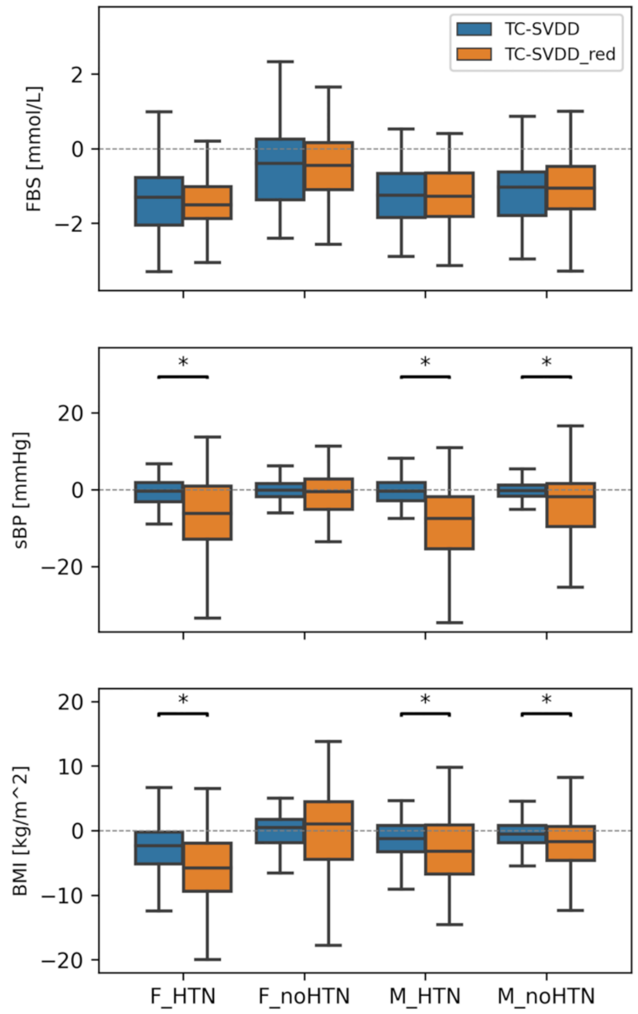
In collaboration with:

University of Toronto
Toronto Metropolitan University
Consiglio Nazionale delle Ricerche

- Lenatti M., Carlevaro A., Keshavjee K., Guergachi A., Mongelli M., Paglialonga A. "*A novel method to derive personalized minimum viable recommendations for type 2 diabetes prevention based on counterfactual explanations*," PLOS ONE, vol. 17(11): e0272825, 2022. https://doi.org/10.1371/journal.pone.0272825
- Carlevaro A., Lenatti M., Paglialonga A., Mongelli M., "*Multi-class counterfactual explanations using support vector data description,*" *IEEE Transactions on Artificial Intelligence*, 2° revision submitted on Oct 03, 2023. Preprint available at: https://doi.org/10.36227/techrxiv.22221007.v1

DATA-DRIVEN MODELS: COUNTERFACTUAL EXPLANATIONS





NEXT STEPS

- multi-class modeling (healthy-preT2D-T2D)
- Personalized definition of controllable features (e.g., if drug.resistant)
- Investigation of causal models
- Characterization of changes as a function of risk

Consiglio Nazionale delle Ricerche

DATA-DRIVEN MODELS: PATIENT TRAJECTORIES

GOAL: monitor the risk of developing T2D based on routinely collected biomarkers up to 8 years before the onset

19 Inputs (t1):	7 Outputs (t2):
• Medications (categorical) • Comorbidities (categorical) • Biomarkers • BMI • Age	• Biomarkers

Algorithm: Multi-input multi-output gaussian process regression model

(adapted from: Chen, Z., B. Wang, and A. N. Gorban. "Multivariate gaussian and student− t process regression for multi-output prediction, 2017")

$$k_{SE}(x, x') = s_f^2 \exp\left(-\frac{\|x - x'\|^2}{2\ell^2}\right)$$

Isotropic kernel

- 5 fold cross validation on training data (80%) : isotropic kernel choice
- Hyperparameter optimization: 28 output covariance matrix hyperparameters + 2 isotropic input kernel hyperparameters + 1 noise variance hyperparameter
- Computation of predictive distribution on testing data (20%)

- Simeone D., Lenatti M., Lagoa C., Keshavjee K, Guergachi A., Dabbene F., Paglialonga A., "*Multi-Input Multi-Output Dynamic Modelling of Type 2 Diabetes Progression*," Studies in Health Technology and Informatics, Proceedings of the European Federation of Medical Informatics Special Topic Conference (EFMI STC 2023), Oct 25-27, 2023, Torino, Italy. In press.

In collaboration with:

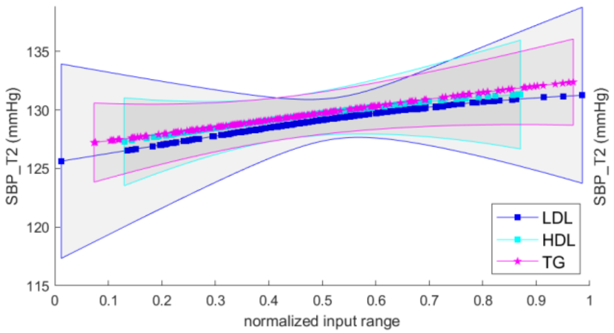
PennState College of Engineering

Toronto Metropolitan University

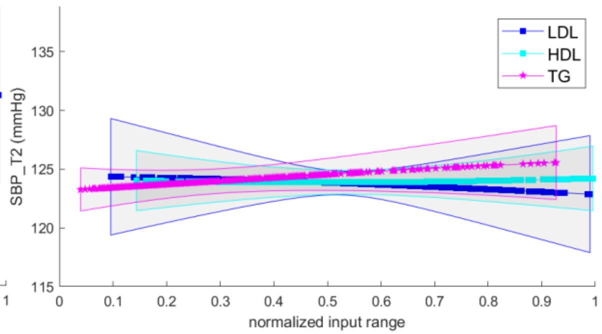
Consiglio Nazionale delle Ricerche

DATA-DRIVEN MODELS: PATIENT TRAJECTORIES

Diabetic (D) group



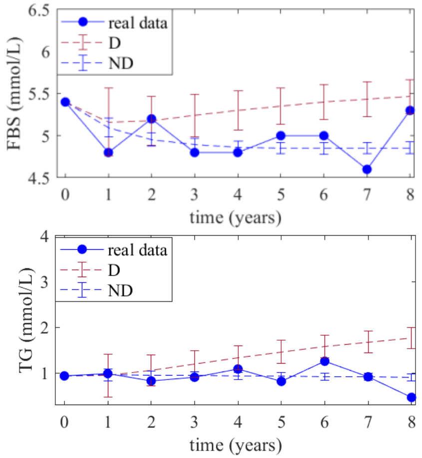
Nondiabetic (ND) group



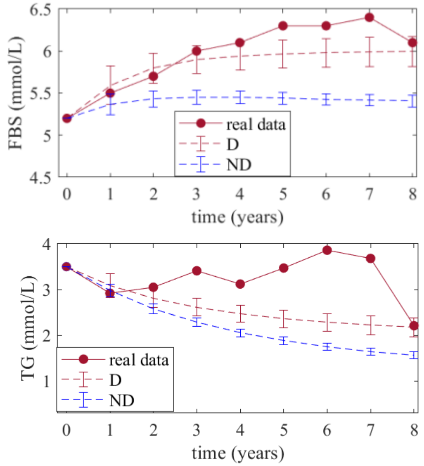
Example of findings from D and ND groups and from sub-groups:

- Stronger positive association between lipids and sBP for D patients compared to ND
- Stronger positive association between age and systolic blood pressure for females compared to males
- Stronger positive association between LDL and sBP for D smokers compared to D nonsmokers
- Stronger negative association between LDL and HDL for D smokers compared to D nonsmokers, for both males and females
- …

Example: patient at low risk of T2D



Example: patient at high risk of T2D



NEXT STEPS

- Analysis and validation of models from sub-groups (vs. sex, risk factors, medications)
- Analysis of different time windows w.r.t. possible onset
- Development of several MIMO models using sub-sets of input/outputs
- Modeling intervention (e.g. physical activity)

Consiglio Nazionale delle Ricerche

Consiglio Nazionale delle Ricerche

BUILD-IT2023 Workshop

October 19-20 2023, Rome

Thank you for your attention

Cnr- Istituto** per le **Applicazioni** del **Calcolo (CNR-IAC)

contact: paolo.tieri@cnr.it

Cnr-Istituto** di **Elettronica** e di **Ingegneria** dell'**Informazione** e delle T**elecomunicazioni (CNR-IEIIT)

contact: alessia.paglialonga@cnr.it

Follow us on: @CNRIAC @IeiitC @cnr.iac @ieiit_cnr