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, *and* Fabrizio Dabbene
Cnr-Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni (CNR-IEIIT), Italy
e-mail: alessia.paglialonga@cnr.it

Filippo Castiglione
Technology Innovation Institute, Biotech Research Center, Abu Dhabi, UAE and Cnr-Istituto per le Applicazioni del Calcolo (CNR-IAC), Rome, Italy

Maria Concetta Palumbo, Paola Stolfi, *and* Paolo Tieri
Cnr-Istituto per le Applicazioni del Calcolo (CNR-IAC), Rome, Italy

This contribution outlines current research aimed at developing models for personalized type 2 diabetes mellitus (T2D) prevention in the framework of the European project PRAESIIDIUM (Physics Informed Machine Learning-Based Prediction and Reversion of Impaired Fasting Glucose Management) aimed at building a digital twin for preventing T2D in patients at risk. Specifically, the modelling approaches include both a multiscale, hybrid computational model of the human metaflammatory (metabolic and inflammatory) status, and data-driven models of the risk of developing T2D able to generate personalized recommendations for mitigating the individual risk. The prediction algorithm will draw on a rich set of information for training, derived from prior clinical data, the individual's family history, and prospective clinical trials including clinical variables, wearable sensors, and a tracking mobile app (for diet, physical activity, and lifestyle). The models developed within the project will be the basis for building a platform for healthcare professionals and patients to estimate and monitor the individual risk of T2D in real time, thus potentially supporting personalized prevention and patient engagement.

Keywords: *counterfactuals, multiscale models, physics-informed machine learning, prediabetes, type 2 diabetes*

1. Introduction

Diabetes is a chronic disease affecting about 10.5% of the global population and characterized by abnormally increased blood glucose level and associated with complications such as cardiovascular and neurological diseases. Type-2 diabetes (T2D) is characterised by insulin resistance (i.e., the insulin hormone cannot effectively regulate the blood glucose levels) and/or a failure in compensatory mechanisms for insulin secretion (i.e., the body does not correctly use the insulin produced) [1]. Prediabetes is an early-stage condition in the "healthy-to-T2D transition" but, unlike T2D, it can be reversed, for example by lifestyle modification. However, despite high-quality prevention guidelines, prevention of T2D remains a global challenge. Developing prevention strategies to reduce the risk of T2D is of paramount importance to limit the burden of disease. Individualized recommendations can be more effective than 'one size fits all' approaches in supporting patient motivation and behaviour change.

To support personalized prevention in patients at risk of T2D, data-driven approaches using digital twin technology can be helpful to support patient awareness and engagement. For example, digital twins can be used to monitor the individual risk in real time, thus helping the patient understand how the risk decreases as a function of preventive strategies (e.g., diet, physical activity, improved lifestyle, medications). Recently, the European project PRAESIIDIUM[1] (Physics Informed Machine Learning-Based Prediction and Reversion of Impaired Fasting Glu-

[1] https://praesiidium.spindoxlabs.com/

cose Management), including 11 participants from seven countries, was launched with the objective to develop an AI-based tool coupled with multi-scale, multi-organ integrated mathematical equations for the real-time prediction of the prediabetes risk of an individual and the identification of personalized recommendations to reduce and monitor the individual risk in real time. Within the framework of PRAESIIDIUM, we have developed a combination of methods, including multiscale modelling and data-driven modelling, as a basis for a digital twin able to characterize the individual risk of developing T2D and identify personalized countermeasures to reduce the risk.

2. Multiscale modelling: Mission-T2D (MT2D)

MT2D is a computational model of the human metabolic and inflammatory status that is determined by the individual dietary and activity habits[2]. MT2D embraces, to a certain degree of sophistication, different levels of description from the molecular/cellular to organs and the whole-body and is based on the sub-jects-specific features to achieve greater generalization and user-customization.

It includes a model for the food intake, stomach emptying and gut absorption of a mixed meal [2] (with the three macronutrients proteins, carbohydrates, fats), a component to account for the effects of physical activity on the hormones' regulation [3], a description of the secretion of Interleukine-6 both by the skeletal muscle during physical activity [4] and by the adipose tissue in resting conditions [5], giving evidence of its dual nature as an adipokine (i.e., adipose tissue-derived cytokine) and as a myokine (i.e., muscle-derived cytokine), and a characterization of energy intake-expenditure balance leading to gaining/losing weight and finally the detailing of the immunological scenario of the subject [6]. All these components are merged into a single, integrated simulation tool with the aim to provide and explore the systemic picture of the metaflammatory status of an individual, that can be potentially exploited proactively to prevent the onset of T2D.

3. Data-driven modelling: Counterfactual explanations and multi-input multi-output dynamic models

Counterfactual explanations (CE) are a local eXplainable AI technique and are defined as the set of minimal changes that, applied to the input features related to a specific instance, can change its predicted class. We developed a new method for generating CE using a Support Vector Data Description classifier to define personalized recommendations to reduce the risk of T2D [7, 8, 9]. Using routinely collected biomarkers extracted from a balanced dataset of Electronic Medical Records (EMRs, derived from the Canadian Primary Care Sentinel Surveillance Network, CPCSSN[3]) of 5582 patients at low/high risk of developing T2D, we determined low-T2D-risk regions with varying specificity, and we assessed the related CE using quantitative performance metrics (e.g., availability, validity, actionability) and a qualitative survey administered to expert clinicians. The minimum viable changes implied a significant modification of fasting blood sugar, systolic blood pressure, triglycerides, and high-density lipoprotein to lower the risk of T2D, particularly in hypertensive patients. The surveyed experts were overall satisfied with the proposed method and provided suggestions for building more meaningful recommendations. Future research will focus on a larger set of input features (e.g., medications) and on the extension to multi-class problems [9] (e.g., to model the risk of multiple conditions, or health-prediabetes-T2D transitions).

In addition to CE, dynamic models were developed to investigate the long-term trajectories of biomarkers in patients at high/low risk of developing T2D and assess the individual risk several years before the onset. We have developed a multi-input multi-output approach based on a Multivariate Gaussian Process Model with an autoregressive structure [10]. Using routinely collected data (biomarkers, comorbidities, and medications) extracted from CPCCSN3 EMRs from three or more consecutive years from 667 T2D patients (before diagnosis) and from 25094 no-T2D patients, we observed that the slow dynamics of biomarkers (e.g., blood pressure, body mass index, lipids, fasting blood sugar) were able to model accurately the long-term evolution in real patients and were consistent with the literature. Future research includes developments of more specific models (e.g., for male/females, for patients

[2] https://kraken.iac.rm.cnr.it/T2DM/
[3] http://cpcssn.ca

with/without comorbidities or risk factors), validation on large real-world datasets, and individualized modelling of the short- and long-term effects of physical activity on T2D risk [11].

4. Conclusions

The combination of multiscale and data-driven models, backed by data from real-world retrospective databases and data from ad hoc, prospective clinical trials in the project PRAESIIDIUM can help in building personalized algorithms for estimating, monitoring, and reducing the individual risk of developing T2D. The proposed modelling approach can be the basis for building a digital twin for T2D prevention that can support individuals in gaining awareness of the factors influencing the risk, thus potentially engaging patients in safeguarding their own health.

ACKNOWLEDGEMENT

This work was supported in part by the European Union, Grant Agreement n. 101095672 (PRAESIIDIUM "Physics informed machine learning-based prediction and reversion of impaired fasting glucose management", call HORIZON-HLTH-2022-STAYHLTH-02). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them. M. Lenatti is a PhD student enrolled in the National PhD in Artificial Intelligence, XXXVIII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma. D. Simeone, P.F. De Paola, and A. Carlevaro are PhD students at Politecnico di Milano (XXXVIII cycle), Politecnico di Bari (National PhD Program in Autonomous Systems, XXXVIII cycle), and Università di Genova (XXXVI cycle), respectively.

REFERENCES

1. Magliano, D. J., Boyko, E. J., Atlas, I. D., et al., (2021), What is diabetes? *IDF DIABETES ATLAS [Internet]. 10th edition*, International Diabetes Federation.

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

3. Palumbo, M. C., Morettini, M., Tieri, P., Diele, F., Sacchetti, M. and Castiglione, F. Personalizing physical exercise in a computational model of fuel homeostasis, *PLoS computational biology*, **14** (4), e1006073, (2018).

4. Morettini, M., Palumbo, M. C., Sacchetti, M., Castiglione, F. and Mazza, C. A system model of the effects of exercise on plasma interleukin-6 dynamics in healthy individuals: role of skeletal muscle and adipose tissue, *PloS one*, **12** (7), e0181224, (2017).

5. Prana, V., Tieri, P., Palumbo, M. C., Mancini, E., Castiglione, F., et al. Modeling the effect of high calorie diet on the interplay between adipose tissue, inflammation, and diabetes, *Computational and mathematical methods in medicine*, **2019**, (2019).

6. Bernaschi, M. and Castiglione, F. Design and implementation of an immune system simulator, *Computers in biology and medicine*, **31** (5), 303–331, (2001).

7. Lenatti, M., Carlevaro, A., Guergachi, A., Keshavjee, K., Mongelli, M. and Paglialonga, A. A novel method to derive personalized minimum viable recommendations for type 2 diabetes prevention based on counterfactual explanations, *Plos one*, **17** (11), e0272825, (2022).

8. Carlevaro, A., Lenatti, M., Paglialonga, A. and Mongelli, M. Counterfactual building and evaluation via explainable support vector data description, *IEEE Access*, **10**, 60849–60861, (2022).

9. Carlevaro, A., Lenatti, M., Paglialonga, A. and Mongelli, M. Multi-class counterfactual explanations using support vector data description, (2023).

10. Simeone, D., Lenatti, M., Lagoa, C., Keshavjee, K., Guergachi, A., Dabbene, F. and Paglialonga, A. Multi-input multi-output dynamic modelling of type 2 diabetes progression, *Submitted to the European Federation of Medical Informatics Special Topic Conference (EFMI STC 2023), Oct 25-26, Torino, Italy*, (2023).

11. De Paola, P. F., Paglialonga, A., Palumbo, P., Keshavjee, K., Dabbene, F. and Borri, A. The long-term effects of physical activity on blood glucose regulation: a model to unravel diabetes progression, *IEEE Control Systems Letters*, (2023).