Special Issue #3

Mathematical Models and Artificial Intelligence for the Prevention of Type 2 Diabetes

Alessia Paglialonga, Laura Azzimonti, Paolo Tieri

Within the European project **PRAESIIDIUM**, numerous research groups combine their expertise to develop innovative strategies for the prevention of type 2 diabetes, combining mathematical models, biological simulations and explainable artificial intelligence tools.

The special issue of this issue, once again, is dedicated to the story of three experts involved in the European project **PRAESIIDIUM:**

Alessia Paglialonga Senior Researcher at CNR - Institute of Electronics and Information and Telecommunications Engineering **(CNR-IEIIT)** in Milan. In the **PRAESIIDIUM project**, **CNR-IEIIT** contributes with two main research activities. The first concerns the **explainability of algorithms** (*AI explainability*), i.e. the adoption of both intrinsic and post-hoc explainability techniques to make the decisions of predictive models interpretable to clinicians and patients. The goal is to ensure a balance between **generalization** (applicability to large cohorts, such as datasets of non-at-risk Caucasian populations) and **personalization** (specific recommendations based on individual profiles), to make data-trained models more transparent and understandable even for nonexperts. The second front concerns the development of **simplified mathematical models** to assess the impact of physical activity (duration, intensity, regularity) on the risk of developing diabetes in the long term.

Laura Azzimonti Lecturer and Senior Researcher at **SUPSI-IDSIA,** Lugano Switzerland. Within the project, he develops "physics-informed" machine learning models, which combine mathematical knowledge of metabolism with real clinical data. This approach makes it possible to build more interpretable and generalizable models, which are fundamental in the medical field. In collaboration with the CNR-IAC, SUPSI is working on a faster version of the type 2 diabetes mellitus simulator (T2DM), capable of providing personalized diabetes risk predictions at low computational cost, which can also be integrated into an application.

Paolo Tieri Senior Researcher at the **CNR-IAC – Institute for Computing Applications** and adjunct professor of **Bioinformatics and Network Medicine** at the Sapienza University of Rome. In the **PRAESIIDIUM** project, CNR-IAC provides synthetic data from an integrated model of metabolism and the human immune system. This model, calibrated on specific characteristics of the patient (diet, physical activity, immunological parameters), allows to simulate the evolution of the metabolic state both over short periods (days) and over prolonged periods (years). The custom data generated in this way feeds machine learning and deep learning models developed by partners, enabling more accurate and adaptable predictions. These models aim to describe the organism in a multi-level systemic perspective, progressively approaching simulations that include cellular, tissue and organ scales.

Integration of mathematical models and clinical data for the personalization of diabetes risk

Our task as **SUPSI-IDSIA,** explains **Laura Azzimonti**, is to integrate patients' clinical and individual data with the expert knowledge contained in the MT2D model developed by the **CNR** (coordinated by **Paolo Tieri). A concrete example is the evaluation of the impact of physical activity on the risk of developing type 2 diabetes.**

The models we develop are designed to describe the evolution of different metabolites over both short and long-time scales, down to the detail of changes induced by a single meal or exercise session. By integrating these models with the data collected by the project, we can enrich datasets limited to specific populations with additional information, thus improving the ability to extrapolate and customize. In this way, the models do not remain anchored to an "average patient” but are adapted to the characteristics of the individual.

The PRAESIIDIUM **approach** is based precisely on this integration of heterogeneous sources: not a single model or method, but a combination of strategies and techniques for data generation and processing. For modelers, the challenge is twofold: on the one hand, to work on population models; on the other, to integrate information relating to the individual patient (eating habits, level of physical activity, clinical parameters). The path is gradual: it starts with the data considered essential by clinicians, and then extends and generalizes when possible, or deepens the detail where necessary to better describe specific processes.

The ultimate goal is to develop a **personalized, adaptable evaluation and forecasting system always under the supervision of the physician**, who remains the ultimate decision-maker in the interpretation and clinical application of the results.

Machine learning models

In general, **Alessia Paglialonga** explains, machine learning models are developed in two main phases. The first concerns **the acquisition and selection of data with which to train the model**. For example, a risk prediction model is trained on a cohort of patients and then validated on a different population, to verify its predictive accuracy.

The second phase consists **of applying the model to the data of individual patients**, at different points in time, updating their clinical status. In this process, the forecasts may vary. More complex models, trained on large amounts of longitudinal data, require robust metrics to constantly monitor accuracy and continuous validation on new populations. In this way, the model that evolves over time can change the internal relationships between the variables, returning different predictions for the same patient than those provided in the past.

Alongside these approaches, **mathematical models** that describe the temporal evolution of specific biomarkers linked to the risk of type 2 diabetes make it possible to study changes over time and dynamically update the patient's status.

Together with **Laura Azzimonti (SUPSI)** we conducted a study to estimate personalized physical activity plans with the aim of reducing the risk of type 2 diabetes in simulated patients. The plans were progressively updated based on the clinical evolution observed over time. This approach showed how the risk prevention interval can change dynamically depending on the patient's status and highlighted how the integration of different methods improves the accuracy of predictions and the adaptability of models.

The power of these models, Tieri adds, lies in the ability to simulate a wide range of possible trajectories and scenarios, providing a repertoire of theoretical answers that can support the work of clinicians. The value of **PRAESIIDIUM** lies precisely in the **flexibility of the models, capable of describing different and constantly evolving situations.**

The value of models

The accuracy of the models varies depending on the approach. Differential **equations** and, more recently, statistical models of machine learning and deep learning can reach theoretically very high levels: the possibility of estimating numerous immunological and metabolic parameters makes it possible to describe in detail the evolution of the patient's state. On the other hand, the validation of agent models, which try to represent very large systems, at an organism scale, and which require further development, is more complex.

A higher level of accuracy requires multi-level **simulations**, capable of describing processes from the inside of the cell to tissues, organs and the entire organism, as does the MT2D model that integrates an agent model of the immune system with a differential equation model of metabolism. Today this goal seems achievable, even if a lot of work is still needed for the fine tuning and validation of such a complex model.

Validation remains crucial, says **Paglialonga**, especially when introducing new data, as well as interaction with clinicians, to whom the limitations of the model also need to be clarified. Each prediction represents a trade-off between the complexity of the system to be described, the number of equations used, and computational costs. We are still far from a fully individual validation, but the verification of population trends and scientific literature provides a solid point of reference. **In conclusion, there is no single accuracy metric – it depends on the context and goal of the model.**

The importance of data and dialogue with decision-makers

The quality, quantity and continuity of data are essential, but today there is still a lack of systematic and digitized collections capable of integrating with the **electronic medical record,** Paglialonga **points out**. This highlights the urgency of dialogue with decision-makers and stakeholders, so that **large-scale digital mapping** becomes a priority, not only for the population at risk but for the prevention of many chronic diseases. Some European countries already have more advanced infrastructures, **Tieri** points out, capable of feeding predictive models and integrating them into clinical systems as decision support tools. A project like **PRAESIIDIUM** can also help raise awareness of health policy on this front. In the future, we imagine models capable of adapting to every single piece of data in the folders, taking full advantage of this wealth of information.

Relationships with doctors and patients

Today, clinicians are familiar with the use of algorithms, while patients show more heterogeneous attitudes. Some are motivated by knowledge and undertake lifestyle changes; others, on the other hand, perceive information as a source of anxiety: *"If I'm not sick now, why bother? I will only take action when necessary."* This limitation remains even in the presence of very accurate predictive models.

In reality, **model-clinician-patient** collaboration is central. The patient himself provides data that feeds the model, and experience – as emerged with COVID – shows how crucial trust between the population and the health system is. A model only works if citizens are aware of the importance of sharing data, allowing predictive systems to be trained on a wider range of diversity. But this requires careful consideration of the state of **public trust** in medical, health and scientific institutions.

Open Science

Researchers do not operate in isolation in laboratories: the models we develop aim to improve the quality of life. The **transparency** of research is a key principle for us, because sharing – between laboratories, with citizens, with decision-makers – is the engine of innovation.

At **PRAESIIDIUM** this approach becomes concrete: sharing data, feedback, results, impacts and expectations is not only good practice, but the necessary condition for building truly useful and accepted solutions. This is the essence of **open science**: participatory, co-created and shared research.

Contact us:

Contact us for more information

praesiidium.project@spindox.it

https://praesiidium.spindoxlabs.com

PRAESIIDIUM