A Computationally Efficient Deep Learning-Based Surrogate Model of Prediabetes Progression

Lea Multerer
IDSIA
SUPSI
Lugano, Switzerland
lea.multerer@supsi.ch

Stefano Toniolo
IDSIA
SUPSI
Lugano, Switzerland
stefano.toniolo@supsi.ch

Sandra Mitrović
IDSIA
SUPSI
Lugano, Switzerland
sandra.mitrovic@supsi.ch

Maria Concetta Palumbo
Institute for Applied Mathematics
National Research Council of Italy
Rome, Italy
c.palumbo@iac.cnr.it

Alessandro Ravoni
Institute for Applied Mathematics
National Research Council of Italy
Rome, Italy
a.ravoni@iac.cnr.it

Paolo Tieri
Institute for Applied Mathematics
National Research Council of Italy
Rome, Italy
p.tieri@iac.cnr.it

Marco Forgione
IDSIA
SUPSI
Lugano, Switzerland
marco.forgione@supsi.ch

Laura Azzimonti
IDSIA
SUPSI
Lugano, Switzerland
laura.azzimonti@supsi.ch

***Abstract*—Early detection of prediabetes is crucial to preventing its progression to diabetes. Providing individuals with a personalized sense of their risk could improve prevention efforts. While complex mathematical models that simulate metabolic and inflammatory processes offer detailed and patient-specific insights, their computational cost usually makes them impractical for real-time prediction on mobile platforms.**

This work introduces a long short-term memory (LSTM) surrogate for the MT2D model, that simulates the main metabolic and inflammatory processes undergoing the transition to prediabetes. The model is developed using a dataset of 43 669 simulated subjects, each with lifestyle inputs and biomarker outputs over six months. Using 8 time series inputs, the surrogate predicts the dynamics of 11 key metabolic and inflammatory outputs, closely replicating the behaviour of the MT2D model.

After training, the proposed LSTM model reduces computational time from an average of 8.4 hours to 0.1 seconds per simulation, making it suitable for mobile device deployment. The model achieves root mean squared errors on the order of 10^{-2} on scaled data, and shows promise for prediabetes risk assessment by capturing trends in inflammatory biomarkers.

This surrogate model can provide real-time and patient-specific insights into the metabolic health, potentially improving the understanding of prediabetes risk.

***Index Terms*—Surrogate, LSTM, Prediabetes, Risk, Input to Output Prediction, Dynamical System**

I. Introduction

Diabetes is one of the biggest modern public health challenges, with an estimated number of 529 million people being affected worldwide in 2021 [1]. Prediabetes is a condition characterized by elevated blood sugar levels that are not yet high enough to be considered diabetes. It is asymptomatic and reversible, making early detection and intervention crucial to preventing the progression to diabetes [2].

Mathematical models of the metabolic and inflammatory system of the human body can offer insights into the progression towards prediabetes. However, due to their complexity, these models are very computationally expensive and surrogate models present a feasible alternative to get predictions in real-time.

A. Surrogates

A surrogate model is an approximate mathematical model usually adopted when the true model is too computationally expensive. By means of a surrogate, tasks that require multiple simulations, such as predicting outputs at new but similar inputs, or solving optimization problems based on model outputs can be solved.

Various surrogate modeling techniques exist, ranging from basic methods like linear regression or support vector machines to more advanced methods. Survey paper [3] provides an overview of many techniques, while [4] reviews surrogates for groundwater modelling. Among the more advanced methods for surrogate modelling is Gaussian process regression [5], [6], a widely adopted nonparametric Bayesian method. In general, the strength of a Bayesian approach is that it provides a probabilistic prediction, allowing the definition of credible intervals. However, the computational cost increases with the dimension of the prior parameters and the input, often requiring the inversion of a large covariance matrix, which depends on the sample size.

Recently, other machine learning techniques, such as deep learning, have been considered for surrogates. Such models consist of layers of connected nodes with adjustable weights, which are optimized based on an objective function during the learning process. After training the model, i.e. after finding optimal weights, inference can be drawn with limited computational effort. As a result of the recent advancements of deep learning methods [7], [8], various approaches have been proposed to better estimate dynamical systems with such approaches [9]–[13]. For blood glucose forecasting, deep learning architectures such as Long Short-Term Memory (LSTM) [14]–[16], convolutional neural networks and recurrent neural networks [17], or deep transfer techniques [18] have been used. Neural networks can also be extended to physics-informed neural networks [19], where the loss function is extended with a term to penalize deviations from the equations governing the system.

B. The MT2D model

MT2D [20], [21] (from Mission-T2D) is a multi-scale and patient-specific computational model that couples ordinary differential equations with an agent-based part. Its goal is to simulate the main metabolic and inflammatory processes involved in the onset of type 2 diabetes in the absence of familiarity. MT2D models seven tissues and compartments (heart, skeletal muscle, liver, gastrointestinal tract, adipose tissue, brain, and other tissues, connected by the arterial and venous blood circulation). The model is implemented in C and the ODEs are solved numerically with the CVODE library [20], [21].

The patient-specific inputs of MT2D can be customized to reflect a wide spectrum of healthy individuals with varying anthropometric features, meal compositions, and exercise routines. Key features include detailed characterizations of the ingestion of mixed meals, energy balance leading to weight changes, immune responses, and the inclusion of subject-specific traits. MT2D can hence give an idea of the long-term effects of the nutritional and exercise patterns for a subject. Consequently, it can be viewed as a tool to forecast the effects of adhering to different lifestyles.

While the dynamic multivariate outputs of MT2D provide a comprehensive view of the immunological and metabolic changes related to type 2 diabetes, its computational complexity is substantial. On average, running one simulation for six months takes 8.4 hours, with the duration ranging from 1.5 hours to 111.8 hours. This is only feasible on high-performance workstations.

Previous work has led to the development of a surrogate for MT2D by means of a random forest algorithm [22], [23]. The first version [22] focused on mapping 10 different inputs - sex, age, baseline weight, height, training session parameters (number, duration, intensity), and dietary intake (carbohydrates, proteins, fats per meal) — to the final endpoint value of three outputs: Glucose base level, BMI and Tumor Necrosis Factor (TNFa). A subsequent version [23] enhanced the surrogate model by predicting the same outputs at 26 time points (weeks), again using a random forest algorithm. This version assessed model performance via root mean squared error, concluding that the surrogate accurately predicted the dynamics of the simulator's output variables. However, a surrogate model that can reproduce a bigger input to output space and is more flexible in terms of analyzed time intervals, while at the same time being suitable for implementation on a mobile device, is missing. Deep learning methods provide the right framework for this purpose.

C. Goal and Structure

The goal of this work is to develop a surrogate for MT2D that could replicate the model's intricate input to output dynamics over time on a fine time grid, at a fraction of the computational demands. As surrogate, a data-driven approach is chosen. The focus will be on deep learning, more precisely on a LSTM architecture.

The surrogate is designed to run on mobile devices, enabling real-time and personalized predictions for new subjects. By generating detailed time series data, the surrogate predicts a subject's metabolic and inflammatory health over the next six months (the available time course of data). While the data does not extend long enough for some subjects to progress to the prediabetic condition, an increase in the inflammatory biomarkers can be used for prediabetes risk prediction.

The rest of this paper is organized as follows: in Section II, the data available from the simulator are prepared, the surrogate architecture and calibration are described and the performance measures of the surrogate are introduced. In Section III, the results are presented and illustrated for two subjects. This work concludes with a discussion and final remarks (Sections IV and V).

II. Methods

A. Data

A substantial dataset from the MT2D simulator is available, comprising 46 170 simulations executed with varying input parameters and followed by an additional 8 748 simulations generated using the latest version of the simulator, with overlapping input parameters between the two sets. These simulations provided a detailed record of inflammation markers, metabolic outcomes, and anthropometric measures, captured at different time intervals (typically every 10 minutes, 30 minutes, or 8 hours, depending on the quantity) over a span of six months.

1) Inputs: The parameters that were varied in the data generation process are described in Table I.

Input Parameters	Input Format for MT2D
Sex Age BMI_0	3 constants
Duration of PA Intensity of PA Frequency of PA	2 time series, non-zero values at PA times
Carbs Proteins Fats	3 time series, non-zero values at meal times

TABLE I

Overview of the MT2D input parameters (first column) varied in the data generating process and their corresponding data format (second column). BMI_0 refers to the BMI at the start of the simulation (depending on weight and height). The time series encapsulate the physical activity (PA) and meal patterns, where the frequency of training sessions and the meal times are encoded by values of either 0 or specific parameters. This reduces the 9 input parameters to 3 constant inputs and 5 time series inputs for MT2D.

Each parameter was varied across multiple levels. The cardio-respiratory fitness level [24] was parametrized by three parameters, leading to an input of two time series. The input data and selected parameter levels were entirely borrowed from the previously published work [22], [23]. Given the satisfactory performance of the previous surrogate models, the available simulations were considered sufficient to capture the interactions between inputs and outputs. In the previous work, it was also observed that meal composition inputs exhibit correlations with each other [22], [23]. However, for consistency with prior research and literature emphasizing the breakdown into carbohydrates, proteins, and fats, they are treated as separate variables.

2) Outputs: MT2D produces a multitude of outputs. In this exploration, outputs for the surrogate were chosen based on their predictive relevance for prediabetes risk (informed by the previous surrogates [22], [23] and expert opinion) and their practicality for general health assessments. A complete list of the outputs that were included in surrogate can be found in Table II.

Outputs	Category
BMI Glucose base level (GBL)	SMB
Glucose Insulin Triglycerides Free fatty acids (FFA)	FMB
Leptin (LEP) Interleukin-6 (IL-6) Interleukin-2 (IL-2) Interleukin-10 (IL-10) Tumor necrosis factor (TNFa)	IB

TABLE II
OUTPUTS OF MT2D INCLUDED IN THE SURROGATE MODEL. ABBREVIATIONS SMB (SLOW METABOLIC BIOMARKERS), FMB (FLUCTUATING METABOLIC BIOMARKERS), AND IB (INFLAMMATORY BIOMARKERS) CLASSIFY THE TIME SERIES DATA BASED ON THEIR TYPE AND THE BEHAVIOR.

The chosen outputs can be categorized based on their type and behaviour, providing insight into how each output behaves over time in response to varying inputs, as follows:

- SMB (Slow Metabolic Biomarkers) - exhibiting slow dynamics over time, such as constants or monotonic changes;
- FMB (Fluctuating Metabolic Biomarkers) - showing daily and weekly fluctuations influenced by variations in diet and physical activity;
- IB (Inflammatory Biomarkers) - either remaining constant or displaying rapid growth over several months with a seasonal pattern. It was hypothesized that increases in these biomarkers could indicate high risk of developing prediabetes.

3) Data Reduction: The output data from MT2D was condensed over time, seeking a balance between capturing long-term trends and short-term dynamics. Outputs of interest were sampled at 8-hour intervals (8am, 4pm, 12pm). To align with this schedule, meal times were adjusted as follows: breakfast was scheduled at 8am, dinner at 4pm, and lunch was divided such that 3/8 of the intake was allocated to 8am and 5/8 to 4pm, with no intake at 12pm to reflect fasting. Furthermore, the start of physical activity was adjusted to the nearest time on this schedule (8am, 4pm, or 12pm) and all simulations were truncated to 540 time steps, corresponding to approximately 180 days, ensuring consistency despite slight variations observed in MT2D simulations. Simulations that terminated early (i.e., stopped before reaching the planned number of time steps) and those with an initial weight of 45kg or less were excluded due to observed unpredictable behavior. Finally, the two datasets mentioned were merged. For each newer simulation, the most similar old simulation was identified by comparing their inputs (considering slight decimal discrepancies) and was replaced with the new simulation.

4) Training Preparation: The final dataset has dimension $[N, T, \Omega]$, where $N = 43669$ represents the number of MT2D simulations, each containing $T = 540$ time steps across $\Omega = 19$ time series. These time series correspond to the 8 inputs listed in the second column of Table I, with the 3 constant inputs also stored in time series format, and the 11 outputs listed in Table II.

The subjects were randomly split into training and testing sets using a 80:20 ratio. After splitting the data, a logarithmic transformation of the form $\overline{y} = \log(y + 1)$ was applied to the IB-outputs since some outputs were growing very fast. Subsequently a Min-Max scaling was performed for all inputs and outputs to confine all quantities within the range $[0, 1]$.

B. Surrogate

As surrogate, a nonlinear mapping between the time series inputs and their corresponding outputs is envisaged by means of a neural network. This Input to Output (I2O) prediction is designed to closely replicate the behavior of the MT2D simulator over time. Given the diverse behaviors of the outputs, it is crucial to carefully assess and select the appropriate neural network architecture, since certain architectures are better suited for predicting specific types of data than others. The primary focus of this work was on the LSTM architecture [25]. LSTMs are a type of recurrent neural network specifically designed for modeling sequential data. They excel at capturing long-term dependencies within sequences, making them well-suited for this task. The concept of the surrogate and its relation to the data is visualized in Figure 1. In the reporting of this work, standard protocols for the development of neural network surrogates [26], [27] have been followed.

C. Implementation and Calibration

The LSTM surrogate was implemented in Python, using PyTorch [28]. All MT2D simulations were used, 20% were reserved for testing. A fixed seed was used to select the split. 8 inputs were processed, as listed in Table I. The inputs sex, age and BMI_0 are static but were treated as constant time series. 11 outputs were processed, as listed in Table II.

Various configurations were explored to optimize the performance for all three types of outputs (SMB, FMB and IB). After an extensive hyperparameter tuning process, based on a cross-validated grid search strategy using Ray [29], focusing on finding the best number of layers, neurons and learning rate, the following setting was chosen:

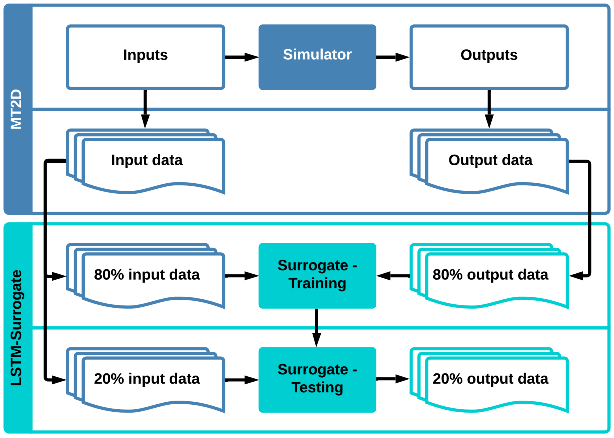

Fig. 1. Visualization of the Input to Output (I2O) approach. The available data from the MT2D simulator is split into training and testing sets. The neural network surrogate's weights are optimized during the training phase and the model is subsequently tested on the remaining data.

- The LSTM was set up with 2 layers, each with 25 neurons.
- Optimization was performed with ADAM, with a learning rate of 0.001 and a weight decay of 10^{-5}.
- A batch size of 512 was chosen.

The smallest mean squared error on the training set was found after 1376 epochs. The training was performed on a computing server equipped with a 64 cores AMD EPYC 7763 and one Nvidia GeForce RTX 4090.

D. Performance evaluation

Let $Y_{n,t,\omega}$ represent the value of the ω-th feature in the n-th simulation at the t-th time step and $\tilde{Y}_{n,t,\omega}$ its approximation. The global performance for each feature ω was measured in terms of the root mean squared error (RMSE),

$$\mathrm{RMSE}_\omega = \sqrt{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=10}^{T}\left(Y_{n,t,\omega} - \tilde{Y}_{n,t,\omega}\right)^2},$$

and the R^2 score,

$$R^2_\omega = 1 - \frac{\sum_{n=1}^{N}\sum_{t=10}^{T}\left(Y_{n,t,\omega} - \tilde{Y}_{n,t,\omega}\right)^2}{\sum_{n=1}^{N}\sum_{t=10}^{T}\left(Y_{n,t,\omega} - \bar{Y}_\omega\right)^2},$$

where $\bar{Y}_\omega$ is the mean of the simulated outputs over time:

$$\bar{Y}_\omega = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=10}^{T} Y_{n,t,\omega}.$$

The R^2 score measures the proportion of the variance in the outputs that are explained by the model. It should be as close to 1 as possible, meaning that the model explains most of the variations in the data. To assess the error over time, the time-specific RMSE ($\mathrm{RMSE}_{t,\omega}$) was considered, which does not include summation over time. Note that the summation over time starts at $t = 10$, since a warm-up period of 3 days was excluded for the performance evaluation. However, these initial time steps are still included in the surrogate.

III. Results

A. MT2D Surrogate

Overall, the total training time of the model was ≈18 minutes, together with around ≈16GB of memory usage dedicated to storing the entire dataset (≈4GB), the model (≈35KB) and all the necessarily computational elements. After training, one model evaluation takes around 0.1 seconds.

In Table III, the RMSE_ω and R^2_ω on the scaled testing data are reported.

Category	**Output** ω	RMSE_ω	R^2_ω
SMB	BMI	0.023	0.98
	GBL	0.012	0.86
FMB	Glucose	0.011	0.93
	Insulin	0.046	0.92
	Triglycerides	0.013	0.85
	FFA	0.015	0.72
IB	Leptin	0.075	0.82
	IL-6	0.058	0.64
	IL-2	0.090	0.82
	IL-10	0.063	0.77
	TNFa	0.057	0.88

TABLE III
ROOT MEAN SQUARED ERROR (RMSE_ω) AND R^2_ω SCORE FOR THE INPUT TO OUTPUT APPROACH FOR EACH FEATURE ω. THESE QUANTITIES ARE CALCULATED ON THE SCALED DATA IN ORDER TO BE COMPARABLE.

All the reported root mean squared errors are in the order of magnitude 10^{-2} for the scaled data. This is higher compared to the random forest surrogates, which achieved a very low root mean squared error. However, this approach is inherently more complex since it maps the inputs to a larger number of outputs across the entire time interval. The RMSE_ω is low for the SMB quantities, together with a reasonably high R^2_ω score. For the FMB quantities, all quantities except Insulin have a low error. The IB quantities are more difficult to estimate, reflected in a higher RMSE_ω. However, their R^2_ω score is still around 0.8 for all quantities, apart from IL-6 with a lower R^2_ω score of 0.64.

The $\mathrm{RMSE}_{t,\omega}$ of the test set over time is visualized in Figure 2. As expected, the error slightly increases with time for all quantities. The error of Insulin fluctuates over time, indicating that the surrogate failed to estimate the behaviour of insulin on a granular level. However, the high R^2_ω score in Table III indicates that the surrogate is able to capture the trend of Insulin.

These results indicate that using a single model for all quantities offers both advantages and limitations. While the surrogate model effectively replicates all 11 quantities, it struggles to capture the fluctuations at a granular level. Performance for individual quantities can be improved by excluding some outputs and focusing hypertuning tuning specifically on the target quantities.

B. Illustration

The results of the surrogate in the previous subsection are based on the 20% of input data that was reserved for testing

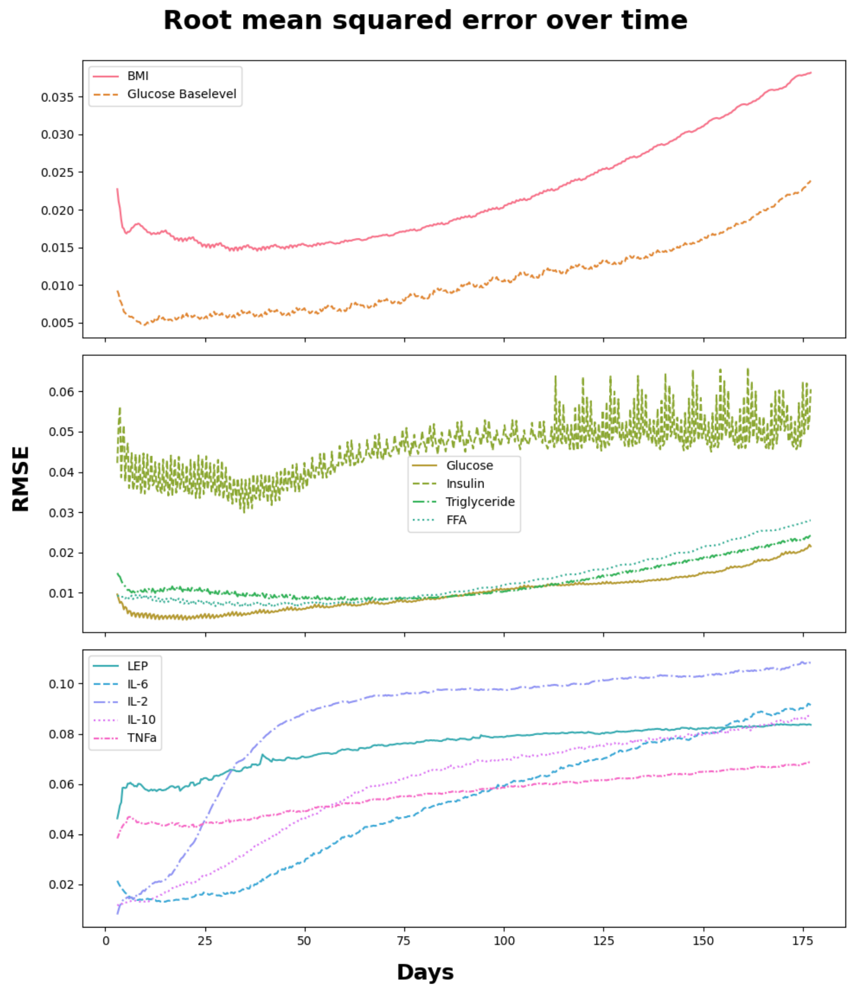

Fig. 2. Root mean squared error over time ($\mathrm{RMSE}_{t,\omega}$).

and not used during the training phase (see Figure 1). For these subjects, the trained model (i.e., the selected neural network configuration with the estimated weights) generates a prediction, given a new patient profile, described in terms of specific diet and physical activity patterns, along with characteristics like sex, age, weight, and height. These predictions are based on the 80% of the data the model learned from during training. These predictions are supposed to mimic the outcomes that would be produced by running the MT2D simulator for the same configuration, but at a fraction of the resources of time and memory.

The performance of the surrogate is illustrated for two subjects of the MT2D test set in Figure 3. Both subjects experienced increases in the inflammatory biomarkers, well captured by the surrogate. For subject 1, glucose baselevel remains constant while for subject 2, glucose baselevel increases, also captured by the surrogate. The FMB quantities are not illustrated in Figure 3, because they exhibit very large daily fluctuations.

IV. Discussion

The surrogate presented in this work is a black-box model, meaning that the output-generating process in not easy to interpret. This stands in contrast to the MT2D simulator, which is a system of ordinary differential equations with a clear physical meaning. Although more explainable strategies for an MT2D surrogate have been explored by means of a Random Forest [22], [23], they did not fully capture the entire output spectrum. Given the primary objective of creating a low-cost surrogate capable of replicating outputs over time, a deep neural network was deemed as the most suitable approach for this task.

On average, a single MT2D simulation required 8.4 hours, while the surrogate, once trained, required only 0.1 seconds on a local workstation, demonstrating that it is suitable for deployment on a mobile device. However, a direct comparison of the trained surrogate with MT2D is very difficult, since they are inherently different. Both the number of time steps and outputs for the surrogate were reduced and a significantly different approach was employed. In terms of precision, the

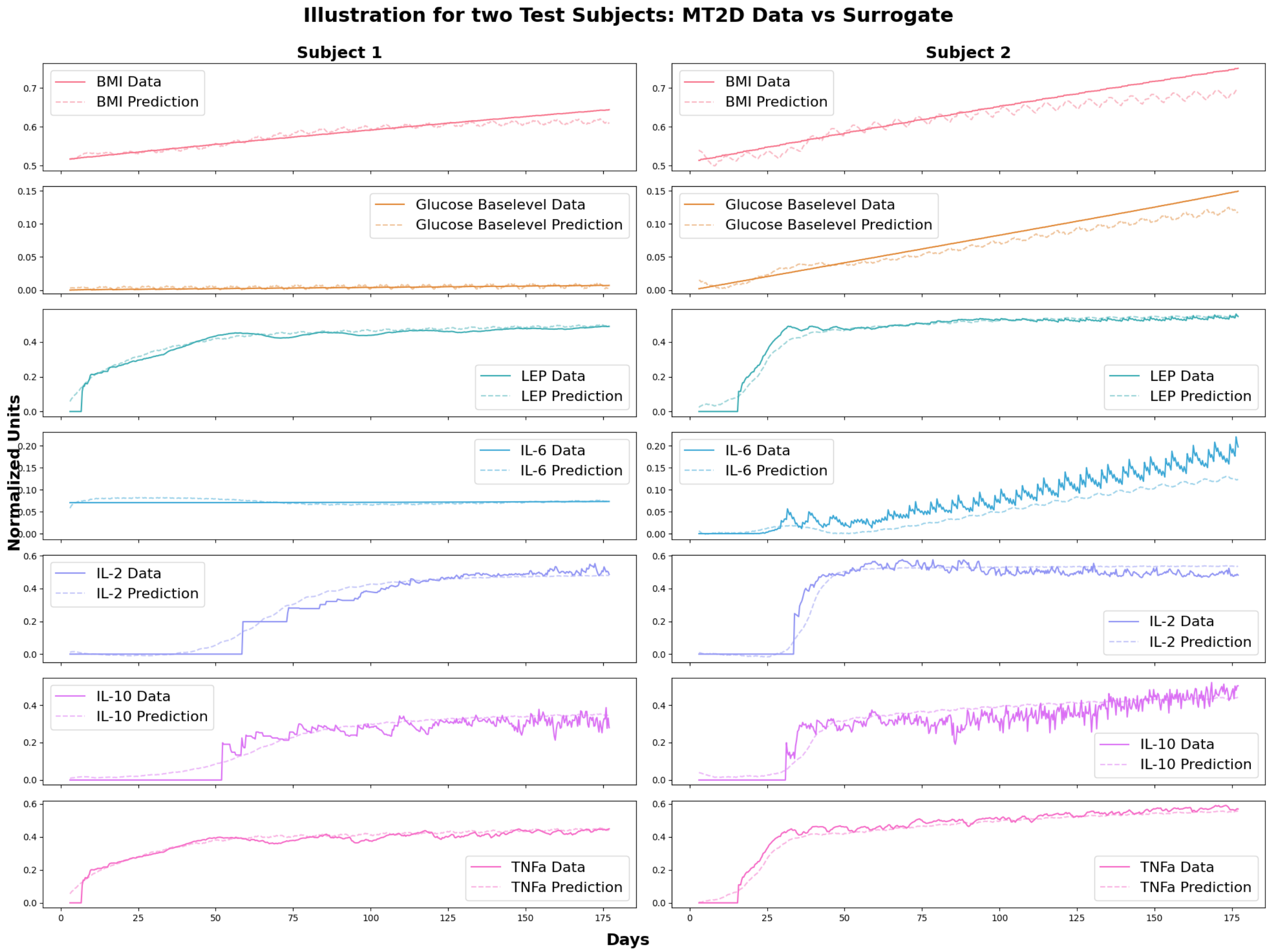

Fig. 3. Illustration of the SMB and IB outcomes of the surrogate for two subjects. Both subjects experienced increases in the IB quantities, Subject 2 with high fluctuations. The surrogate captures the underlying trends in the data, but does not capture all the fluctuations. The FMB quantities are not illustrated, since they fluctuate heavily over 6 months.

surrogate reported root mean squared errors in the magnitude of 10^{-2} on the scaled data.

A single surrogate model was chosen to predict all 11 outputs from the 8 inputs. While it occasionally struggled to capture fluctuations in the data, it accurately captured the overall trends. A surrogate that predicts all outputs simultaneously can profit from all the information available, but may be less tailored to specific quantities (e.g. insulin).

The inflammatory biomarkers (see IB in Table II and the five lower plots of Figures 3) exhibit approximately a logistic growth pattern. This suggests that a logistic growth model could be fitted to the trajectories of the surrogate, summarizing the growth pattern to one feature. These simple metrics can then be used to categorize subjects into different risk categories for developing an unhealthy condition, such as prediabetes. This approach should be quite stable to fluctuations in the data.

V. Conclusion

In this work, a deep learning surrogate was developed for the MT2D simulator, a complex and detailed but computationally expensive model that describes the human metabolic and inflammatory system. The available data was reduced in terms of time resolution and number of output channels. A LSTM architecture was chosen to predict outputs from inputs over time, closely replicating the behavior of the MT2D simulator. Specifically, the LSTM was designed to accept 8 time series inputs and generate 11 relevant MT2D outputs. The architecture yielding the smallest root mean squared error (RMSE) on all features was found to have 2 layers, each layer comprising of 25 neurons.

The proposed surrogate is suitable for mobile devices and can generate patient-specific profiles informed by the MT2D data. Furthermore, the surrogate offers potential for prediabetes risk assessment based on the prediction of elevated inflammatory biomarkers.

Several potential extensions of this work can be considered, including exploring different time intervals for the Input to Output approach. For instance, moving from 8-hour time steps to 1-day intervals could create a coarser surrogate, while refining time steps may capture seasonal dynamics more effectively. Additionally, quantifying the uncertainty of the predictions using techniques such as Bayesian neural networks would be an interesting extension.

Acknowledgment

This work was supported in part by the European Union and by the Swiss State Secretariat for Education, Research and Innovation (SERI) through project PRAESIIDIUM "Physics informed machine learning-based prediction and reversion of impaired fasting glucose management" under Grant 101095672. 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.

References

[1] G. . D. Collaborators, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021," *The Lancet*, vol. 402, no. 10397, pp. 203–234, 2023.

[2] M. R. Rooney, M. Fang, K. Ogurtsova, B. Ozkan, J. B. Echouffo-Tcheugui, E. J. Boyko, D. J. Magliano, and E. Selvin, "Global Prevalence of Prediabetes," *Diabetes Care*, vol. 46, no. 7, pp. 1388–1394, Jul. 2023. [Online]. Available: https://diabetesjournals.org/care/article/46/7/1388/148937/Global-Prevalence-of-Prediabetes

[3] A. Bhosekar and M. Ierapetritou, "Advances in surrogate based modeling, feasibility analysis, and optimization: A review," *Computers & Chemical Engineering*, vol. 108, pp. 250–267, Jan. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0098135417303228

[4] M. J. Asher, B. F. W. Croke, A. J. Jakeman, and L. J. M. Peeters, "A review of surrogate models and their application to groundwater modeling," *Water Resources Research*, vol. 51, no. 8, pp. 5957–5973, Aug. 2015. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015WR016967

[5] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, issue: 3.

[6] A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorpt, Eds., *Handbook of spatial statistics*. Boca Raton: CRC Press, 2010, oCLC: 652826600.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: https://www.nature.com/articles/nature14539

[8] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608014002135

[9] M. Forgione and D. Piga, "Continuous-time system identification with neural networks: Model structures and fitting criteria," *European Journal of Control*, vol. 59, pp. 69–81, May 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0947358021000169

[10] M. Forgione, F. Pura, and D. Piga, "From System Models to Class Models: An In-Context Learning Paradigm," *IEEE Control Systems Letters*, vol. 7, pp. 3513–3518, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10324309/

[11] N. Geneva and N. Zabaras, "Transformers for modeling physical systems," *Neural Networks*, vol. 146, pp. 272–289, Feb. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608021004500

[12] P. Rajendra and V. Brahmajirao, "Modeling of dynamical systems through deep learning," *Biophysical Reviews*, vol. 12, no. 6, pp. 1311–1320, Dec. 2020. [Online]. Available: http://link.springer.com/10.1007/s12551-020-00776-4

[13] Y. Zhao, C. Jiang, M. A. Vega, M. D. Todd, and Z. Hu, "Surrogate Modeling of Nonlinear Dynamic Systems: A Comparative Study," *Journal of Computing and Information Science in Engineering*, vol. 23, no. 1, p. 011001, Feb. 2023.

[14] S. H. A. Faruqui, Y. Du, R. Meka, A. Alaeddini, C. Li, S. Shirinkam, and J. Wang, "Development of a Deep Learning Model for Dynamic Forecasting of Blood Glucose Level for Type 2 Diabetes Mellitus: Secondary Analysis of a Randomized Controlled Trial," *JMIR mHealth and uHealth*, vol. 7, no. 11, p. e14452, Nov. 2019. [Online]. Available: https://mhealth.jmir.org/2019/11/e14452

[15] S. Mirshekarian, R. Bunescu, C. Marling, and F. Schwartz, "Using LSTMs to learn physiological models of blood glucose behavior," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Seogwipo: IEEE, Jul. 2017, pp. 2887–2891. [Online]. Available: https://ieeexplore.ieee.org/document/8037460/

[16] E. M. Aiello, G. Lisanti, L. Magni, M. Musci, and C. Toffanin, "Therapy-driven Deep Glucose Forecasting," *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103255, Jan. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0952197619302313

[17] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional Recurrent Neural Networks for Glucose Prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 603–613, Feb. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8678399/

[18] X. Yu, T. Yang, J. Lu, Y. Shen, W. Lu, W. Zhu, Y. Bao, H. Li, and J. Zhou, "Deep transfer learning: a novel glucose prediction framework for new subjects with type 2 diabetes," *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 1875–1887, Jun. 2022. [Online]. Available: https://link.springer.com/10.1007/s40747-021-00360-7

[19] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, May 2021. [Online]. Available: https://www.nature.com/articles/s42254-021-00314-5

[20] 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*, vol. 14, no. 4, p. e1006073, Apr. 2018. [Online]. Available: https://dx.plos.org/10.1371/journal.pcbi.1006073

[21] M. C. Palumbo, A. A. De Graaf, M. Morettini, P. Tieri, S. Krishnan, and F. Castiglione, "A computational model of the effects of macronutrients absorption and physical exercise on hormonal regulation and metabolic homeostasis," *Computers in Biology and Medicine*, vol. 163, p. 107158, Sep. 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010482523006236

[22] P. Stolfi, I. Valentini, M. C. Palumbo, P. Tieri, A. Grignolio, and F. Castiglione, "Potential predictors of type-2 diabetes risk: machine learning, synthetic data and wearable health devices," *BMC Bioinformatics*, vol. 21, no. S17, p. 508, Dec. 2020. [Online]. Available: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-03763-4

[23] P. Stolfi and F. Castiglione, "Emulating complex simulations by machine learning methods," *BMC Bioinformatics*, vol. 22, no. S14, p. 483, Nov. 2021. [Online]. Available: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-021-04354-7

[24] V. H. Heyward and A. L. Gibson, *Advanced fitness assessment and exercise prescription*, 7th ed. Champaign, IL: Human Kinetics, 2014.

[25] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: https://direct.mit.edu/neco/article/9/8/1735-1780/6109

[26] H. R. Maier, A. Jain, G. C. Dandy, and K. Sudheer, "Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions," *Environmental Modelling & Software*, vol. 25, no. 8, pp. 891–909, Aug. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364815210000411

[27] W. Wu, G. C. Dandy, and H. R. Maier, "Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling," *Environmental Modelling & Software*, vol. 54, pp. 108–127, Apr. 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364815213003198

[28] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," 2019, version Number: 1. [Online]. Available: https://arxiv.org/abs/1912.01703

[29] P. Moritz, R. Nishihara, S. Wang, A. Tumanov, R. Liaw, E. Liang, M. Elibol, Z. Yang, W. Paul, M. I. Jordan, and I. Stoica, "Ray: A Distributed Framework for Emerging AI Applications," 2017, version Number: 2. [Online]. Available: https://arxiv.org/abs/1712.05889