Evaluation of Quantitative Systems Pharmacology and Machine Learning Models for Blood Glucose Prediction

Objectives

This project has the following three objectives, which would be taken by the project team side-by-side:

1. Writing a white paper on Role of Artificial Intelligence in Healthcare.
2. Literature survey of existing quantitative systems pharmacology (QSP) and machine-learning (ML) models for blood glucose prediction.
3. Evaluating the selected QSP and ML models for blood glucose prediction.
4. Understanding the dynamics of influence propagation inside a physician social network for generating insights into prescribing patterns for identifying rare disease patients

These objectives may evolve over the course of the project as needed by Rx Data Science.

Objective 1: White Paper on Role of Artificial Intelligence in Healthcare

In this article, we will chart a road map on how Artificial Intelligence (AI) has been currently used to solve problems and challenges in different healthcare domains. Also, we will provide certain novel ideas on how existing machine learning and deep learning algorithms (e.g., generative adversarial networks) can be integrated with cognitive algorithms to enable researchers to solve emerging problems in healthcare via AI. Some part of this white paper had already been written and these and other parts will be improved and added respectively.

Objective 2: Literature Survey of Existing Mathematical and Neural Network Models for Blood Glucose Prediction

Diabetes mellitus is a major, and increasing, global problem. The number of people affected with diabetes in 2000 was estimated to be 171 million worldwide. This figure is predicted to rise to 366 million by 2030 [2], which represent around 4.4% of the estimated worldwide population. However, it has been shown that, through good management of blood glucose levels (BGLs), the associated and costly complications can be reduced significantly [1].

The human body requires the maintenance of blood glucose (BG) levels in a very narrow range (70–110 mg/dl) [3]. Many different factors affect these levels. The pancreas releases insulin and glycogen hormones to regulate the BG levels. Type 1 diabetes mellitus (T1D) is the consequence of an autoimmune attack on the pancreas that significantly impairs insulin production. Complications due to T1D include neuropathy, nephropathy, and retinopathy, with diabetes being a leading cause of renal failure, new blindness, and non-trauma amputations. These complications, in turn, impose a significant economic burden to society, with indirect and direct costs estimated at $245 billion for the U.S. alone in 2012 [4]. Even more alarming, diabetes is a rapidly growing global epidemic for which the Centers for Disease Control estimate will affect 1 in 3 adults in the U.S. by 2050 if current trends continue [2]. For Type 1 (insulin-dependent) diabetics, the most common method for management is through monitoring the blood glucose level (BGL) using finger-prick blood tests taken several times a day, and adjusting insulin doses based on these readings. For a dynamic, nonlinear, and complex condition such as diabetes this can be far from satisfactory. Factors such as insulin type and dose, diet, stress, exercise, illness, and pregnancy all have significant influences on the BGL. Management may be compromised through lack of data and, for some patients, an inability to interpret data adequately [1].

Quantitative Systems Pharmacology models have been widely used in biology and medicine

for several decades, and have, in certain areas, already gained widespread acceptance. Few such QSP models are the Automated Insulin Dosage Advisor (AIDA) [5] and the Meal Simulation Model of the Glucose-Insulin System [6]. AIDA models the steady-state glucose-insulin interactions over a 24-hour period by describing the physiology of a person with Type-1 diabetes mellitus. Similarly, the Meal Simulation Model of the Glucose-Insulin System develops a model for the human glucose-insulin response after a meal. For making BGL predictions using these models require a previous understanding of insulin and glucose metabolism. This type of model contains several physiological parameters that need to be set prior to their use to make BG predictions. For example the AIDA model takes into consideration a number of factors: carbohydrate intake (amount in grams and time), insulin (type, amount, and time), kidney function, and insulin sensitivity. Using these variables, a compartmental model of the various interactions within the body is described by a series of differential equations to build a picture of BGL fluctuations over a 24-hour period in time steps of 0.25 hours. It assumes a patient is unable to produce endogenous insulin, as is the case with Type-1 diabetics. [5]. Another such models take as input age, height, weight, average plasma concentration, bone mass, carbohydrate intake, Fat intake, Muscle (grams), physical activity (PA) factor, food intake, body mass index to calculate the insulin resistance [7]. They then solve differential equations to predict the BG levels of the patients.

A recent advent in physiological modeling is multi-level models, i.e. models that simultaneously capture several layers of biological organization. Apart from these types of model that spans several layers of biological organization; physiological models can also capture multiple timescales. There are also models that not only describe the immediate response, but also describe longer-term dynamics, such as the increase in weight as a result of over-eating, longer-term improvements of a drug, as well as disease development and progression [8]. As such a thorough literature review of the current physiological approaches has to be made to better understand the dynamics of physiological modeling for diabetes.

Due to the recent advances in the field of deep learning, there exists an extensive literature in the field of BGL prediction and control using ML models specifically artificial neural networks (ANN). ANNs represent a powerful modeling technique for pattern recognition, time series forecasting, and regression problems [2]. Their modeling ability relies on the fact that they do not require in-depth knowledge of the relationship between intrinsic variables for a particular problem, or the structure relating the variables required for mathematical modeling. Instead, an ANN is trained to recognize patterns in a dataset, and these patterns are effectively stored in the interconnected neuron weights of the ANN. However, inter-subject variability is one of the major challenges to be addressed in any BGL monitoring system (mathematical models) which attempts to control the BGL of a patient, which is why ANNs may be a possible solution. An ANN can be trained on an individual's BGL; it can incorporate the factors specific to that person, without an in-depth knowledge of the individual-specific interactions which link them all. These factors can include, but are not limited to body mass index (BMI), age, gender, pregnancy, carbohydrate intake, exogenous insulin injections, exercise, and stress [17].

In this objective, we aim to perform an extensive study of the mathematical and machine learning based models present in literature and to compare their efficacy in BGL predictions.

Objective 3: Evaluation of Mathematical Models and Neural Network Models for Blood Glucose Prediction

In this objective, we will be generating scenarios using AIDA and Meal Simulation model of the Glucose-Insulin System in Matlab and evaluating these using different Machine learning models.

Methodology for objective 3

BG predictions are mostly done using physiological and data-driven models. As such we will use two types of models in this project for BG prediction:

I) Physiological models particularly Quantitative System Pharmacology (QSP) models
II) Data Driven model particularly machine learning models

Physiological Models

Physiological models take into account the physiological characteristics of the patient to predict the BG levels. In this project we will be using the AIDA and the Meal Simulation model of the Glucose-Insulin System as our baseline models.

AIDA

AIDA incorporates a compartmental model that describes glucose-insulin interaction in people completely lacking endogenous insulin secretion — i.e. insulin-dependent patients with type 1 diabetes mellitus. The AIDA model contains a single extra-cellular glucose compartment into which glucose enters via both absorptions from the intestine and glucose production from the liver. The model also contains separate compartments for plasma and 'active' insulin, the latter being responsible for glycemic control while insulin is removed from the former by liver degradation [5].

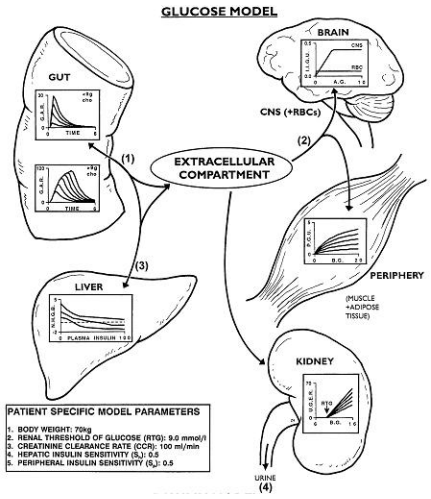

Figure 1: Schematic summarizing the anatomical basis and physiological functions of the AIDA model.

Meal Simulation model of the Glucose-Insulin System [6]

Meal simulation model models the human glucose-insulin response after a meal. This model describes the dynamics of the system using ordinary differential equations. The authors used their model to simulate the glucose-insulin response after one or more meals, for normal human subjects and for human subjects with various kinds of insulin impairments. The impairments were represented as alternate sets of parameter values and initial conditions. This model can be simulated in Matlab using the SimBiology package [12].

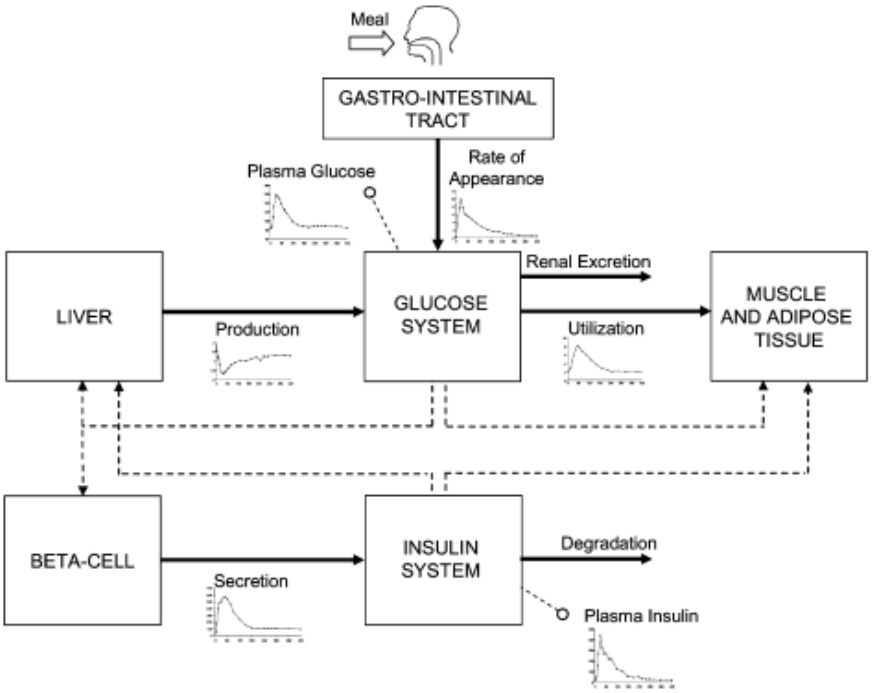

Figure 2: Scheme of the glucose-insulin control system in the Meal Simulation model.

We will also look into other complexities common to disease modeling, such as inter-subject variability (e.g., disease state, genetic and phenotypic differences in metabolism, disease etiology, diet, and therapeutic intervention history) and intra-subject variability (e.g., assay variability, day-to-day compliance with treatment(s), food intake, and exercise).

Data driven Models

Data driven models completely rely only on BG data and other inputs. Data-driven models are typically based on machine learning techniques and use techniques such as genetic algorithms, robust filters, fuzzy logic, rule-based models, multi-model approaches, autoregressive models, random forests, support vector regression, and artificial neural networks models [3]. Deep Neural Networks, especially of the convolutional type (DCNNs), have started a revolution in the field of artificial intelligence and machine learning, triggering a large number of commercial ventures and practical applications. One of these ventures is the prediction of BG level for Diabetic patients. Researchers have used Recurrent ANN and Multi-Layer Perceptron's (MLP) (designed and trained using MATLAB) with to predict future BGL values using input parameters including insulin, diet, exercise, BGL, and other factors [13]. In this project we intent to apply machine learning models such as Convolutional Neural Networks [14], Long Short-Term Memory [15], and Generative Adversarial network (GANs) [16] models. We will generate data and scenarios from AIDA and Meal Simulation model (QSP models). Next, we will calibrate the ML models (statistical, traditional, and deep, including GANs) to the QSP data from AIDA and Meal Simulation model models to make blood glucose predictions for diabetic patients.

Objective 4: Understanding the dynamics of influence propagation inside a physician social network for generating insights into prescribing patterns and identifying rare disease patients

We divide this objective into two sub-parts.

Objective 4.1: Understanding the dynamics of influence propagation inside a physician social network for generating insights into prescribing patterns

Understanding how relationships are structured in physician networks provides insights into how these networks influence physicians' beliefs and behaviors. Even though textbooks, journal, or online articles are important sources of information for many physicians [18], nevertheless physicians mostly rely on colleagues for new information to help them interpret the medical literature and to obtain specific advice about the care of their patients. In addition, many physicians rate colleagues as their most valued source of information [19-21]. As a result, the knowledge and beliefs of physicians and how they share information with each other likely shapes the prescription behavior of physicians. The medications that a physician prescribes may likely be influenced by her interpersonal communication (informal consultations) with the members of her personal network. These consultations may aid in the adoption/diffusion of medications among physicians. The adoption of new medicines spreads may differ from physician-to-physician. While some physicians may adopt it early; others may adopt it late or may not adopt it at all [22]. The diffusion process could be thought of as a "social process" that is highly dependent on the members of the social system. For example, it has been shown that physicians who are "socially proximate" in a social environment often use one another as information sources to manage the uncertainty of adopting new antibiotic drugs (Zheng, 2010). For understanding the diffusion of medication, we first need to understand the dynamics of adoption inside the underlying social network. Multiple models of information diffusion have been introduced in the literature; some of the most widely studied models are: The Linear Threshold Model, the Independent Cascade Model, and the general threshold model [23]. The basic assumption of these diffusion models is that in a social network people tend to perform an action (adopting a medication) if they see their social contacts (friends, family, and acquaintances) performing that action [23]. In this project we would use these diffusion models to understand the dynamics of influence propagation inside a physician social network. We will also use deep learning approaches such as Generative adversarial networks (GAN) and Long short term memory networks (LSTM) to generate and predict future diffusion behavior of physicians. In this project we will try to generate insights into the prescribing patterns for **OnPattro.**

Objective 4.2: Identifying rare disease patients specifically patients suffering from hereditary transthyretin-mediated (hATTR) amyloidosis

Since the past three years, the last day of February has been designated as Rare Disease Day to call attention to public health issues associated with rare diseases, which affect nearly 30 million Americans and countless others around the world [24]. By their nature, rare diseases are difficult to diagnose. In this objective we would create algorithms that extracts patterns from patient journey of patients suffering from rare diseases for easier identification. We will be targeting patients suffering from hereditary transthyretin-mediated (hATTR) amyloidosis.

The rare-disease problem is fundamentally unsupervised in the sense that patients don't have a rare disease label attached to their diagnoses and medicine refilling behavior. Thus, methods used to identify patients and their treating physicians will rely on unsupervised machine learning methods like clustering. The project will start with identifying certain rare disease conditions and generate a framework that could be applied to identifying other rare disease conditions.

Timeline

Tasks	Number of Months										
	0	3	6	9	12	15	18	21	24	27	30
Completion of white paper on AI in Healthcare											
Literature review and problem identification of Type 1 Diabetes											
Generation of scenarious from QSP models											
Development of ML models on scenarios from QSP models											
Improvement of ML models on scenarios from QSP models											
Understanding the dynamics of diffusion inside a physician social network											
Predicting diffusion inside a physician social network											
Development of algorithms for finding rare disease patients											
Integration of development algorithms and reporting of outcomes											

Budget

Cost Head	Amount (INR)
Number of months (1)	30
Number of months for PG students (1.1)	10
Number of months for Graduate students (1.2)	30
Number of Post Graduate(PG) students (2)	2
Number of Gradute students (2.1)	1
Stipend per PG student per month (3)	35000
Stipend per Graduate student per month (3.1)	12400
Number of faculty (4)	1
Total student stipend (5) = ((1.1) *(2) * (3))+((1.2)*(2.1)*(3.1))	1,072,000
Consultancy fee for faculty into his PDA (6)	150,000
Consumables (7)	50,000
Total cost (8) (= (5)+(6)+(7))	1,272,000
Institute overhead (@ 20%) (9)	254,400
Total cost after Institute overhead (10) (= (8) + (9))	1,526,400
Total cost after Institute overhead (= (11) * 0.015 USD/INR)	22,896
	USD 22896

References

1. Robertson, G., Lehmann, E. D., Sandham, W., & Hamilton, D. (2011). Blood glucose prediction using artificial neural networks trained with the AIDA diabetes simulator: a proof-of-concept pilot study. Journal of Electrical and Computer Engineering, 2011, 2.
2. S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, “Global prevalence of diabetes: estimates for the year 2000 and projections for 2030,” Diabetes Care, vol. 27, no. 5, pp. 1047–1053, 2004.
3. Contreras, I., Oviedo, S., Vettoretti, M., Visentin, R., & Vehí, J. (2017). Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models. PloS one, 12(11), e0187754.
4. Cobelli C, Renard E, Kovatchev B. Artificial pancreas: Past, present, future. Diabetes. 2011;60: 2672–2682. pmid:22025773
5. E. D. Lehmann and T. Deutsch, “A physiological model of glucose-insulin interaction in type 1 diabetes mellitus,” Journal of Biomedical Engineering, vol. 14, no. 3, pp. 235–242, 1992.
6. Dalla Man, C., Rizza, R. A., & Cobelli, C. (2007). Meal simulation model of the glucose-

insulin system. IEEE Transactions on biomedical engineering, 54(10), 1740-1749.
7. Brännmark, C., Nyman, E., Fagerholm, S., Bergenholm, L., Ekstrand, E. M., Cedersund, G., & Strålfors, P. (2013). Insulin signaling in type 2 diabetes experimental and modeling analyses reveal mechanisms of insulin resistance in human adipocytes. Journal of Biological Chemistry, 288(14), 9867-9880.
8. Nyman, E., Rozendaal, Y. J., Helmlinger, G., Hamrén, B., Kjellsson, M. C., Strålfors, P., ... & Cedersund, G. (2016). Requirements for multi-level systems pharmacology models to reach end-usage: the case of type 2 diabetes. Interface Focus, 6(2), 20150075.
9. Cobelli C, Renard E, Kovatchev B. Artificial pancreas: Past, present, future. Diabetes. 2011; 60: 2672–2682. https://doi.org/10.2337/db11-0654 PMID: 22025773
10. Hinshaw L, Dalla Man C, Nandy DK, Saad A, Bharucha AE, Levine JA, et al. Diurnal pattern of insulin action in type 1 diabetes: implications for a closed-loop system. Diabetes. American Diabetes Association; 2013; 62: 2223–9. https://doi.org/10.2337/db12-1759 PMID: 23447123
11. Bailey TS, Chang A, Christiansen M. Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm. J Diabetes Sci Technol. Diabetes Technology Society; 2015; 9: 209–14. https://doi.org/10.1177/1932296814559746 PMID: 25370149
12. Matlab (2019), Simulating the Glucose-Insulin Response, https://in.mathworks.com/help/simbio/examples/simulating-the-glucose-insulin-response.html
13. W. Sandham, D. Nikoletou, D. Hamilton, K. Patterson, A. Japp, and C. Macgregor, "Blood glucose prediction for diabetes therapy using a recurrent artificial neural network," in Proceedings of the 9th European Signal Processing Conference (EUSIPCO '98), vol. 11, pp. 673–676, 1998.
14. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
15. Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.
16. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
17. Mhaskar, H. N., Pereverzyev, S. V., & van der Walt, M. D. (2017). A deep learning approach to diabetic blood glucose prediction. Frontiers in Applied Mathematics and Statistics, 3, 14.
18. Haug, J. D. (1997). Physicians' preferences for information sources: a meta-analytic study. Bulletin of the Medical Library Association, 85(3), 223.
19. Gruppen, L. D., Wolf, F. M., Voorhees, C. V., & Stross, J. K. (1987). Information-seeking strategies and differences among primary care physicians. Journal of Continuing Education in the Health Professions, 7(3), 18-26.
20. Williamson, J. W., German, P. S., Weiss, R., Skinner, E. A., & Bowes, F. (1989). Health science information management and continuing education of physicians: a survey of US primary care practitioners and their opinion leaders. Annals of Internal Medicine, 110(2), 151-160.
21. Cullen, R. (1997). The medical specialist: information gateway or gatekeeper for the family practitioner. Bulletin of the Medical Library Association, 85(4), 348.
22. Mahajan, V., Muller, E., & Bass, F. M. (1991). New product diffusion models in marketing: A review and directions for research. In Diffusion of technologies and social behavior (pp. 125-177). Springer, Berlin, Heidelberg.
23. Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 137-146). ACM.
24. https://www.massbio.org/events/event-archive/rare-disease-day-2017-patientinvolvement-in-

research-1806