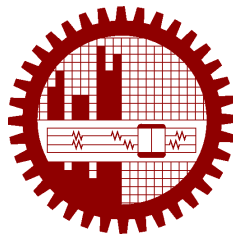


M.Sc. Engg. (CSE) Thesis

Integrated Physiological Model of Virtual Diabetic Patient for Machine Learning Research

Submitted by
Md. Jahirul Islam
1014052074

Supervised by
Dr. Abu Sayed Md. Latiful Hoque



Submitted to
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

in partial fulfillment of the requirements for the degree of
Master of Science in Computer Science and Engineering

March 2021

Candidate's Declaration

I, do, hereby, certify that the work presented in this thesis, titled, “Integrated Physiological Model of Virtual Diabetic Patient for Machine Learning Research”, is the outcome of the investigation and research carried out by me under the supervision of Dr. Abu Sayed Md. Latiful Hoque, Professor, Department of CSE, BUET.


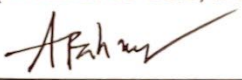

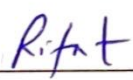

I also declare that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Md. Jahirul Islam

1014052074

The thesis titled “**Integrated Physiological Model of Virtual Diabetic Patient for Machine Learning Research**”, submitted by Md. Jahirul Islam, Student ID 1014052074, Session October 2014, to the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, has been accepted as satisfactory in partial fulfilment of the requirements for the degree of Master of Science in Computer Science and Engineering and approved as to its style and contents on March 29, 2021.

Board of Examiners

1. _____
Dr. Abu Sayed Md. Latiful Hoque
Professor
Department of CSE, BUET, Dhaka

Chairman
(Supervisor)
2. _____
Dr. A. K. M. Ashikur Rahman
Professor and Head
Department of CSE, BUET, Dhaka

Member
(Ex-Officio)
3. _____
Dr. Md. Monirul Islam
Professor
Department of CSE, BUET, Dhaka

Member
4. _____
Dr. Rifat Shahriyar
Associate Professor
Department of CSE, BUET, Dhaka

Member
5. _____
Dr. Md. Adnan Hasan Masud
Consultant
Department of Hematology
BSMMU, Shahbagh, Dhaka (Specialist, Diabetes)

Member
(External)

Acknowledgement

All praise is due to Almighty, the most gracious and the most merciful, who gave us everything for living, thinking, studying and enjoying the life. I would like to express my gratitude to my supervisor, Professor Dr. Abu Sayed Md. Latiful Hoque, for every kind of support during my research. Working with him was a great and invaluable one-time opportunity for me in my life which allowed me to learn how to initiate a solid research, how to conduct it and how to complete it. His wise contributions along with kind encouragements were always supportive in tackling difficulties whenever I faced a disappointing deadlock in my research.

Finally, I give my best thanks to my family and friends for all their support during my M.Sc. studies.

Dhaka
March 29, 2021

Md. Jahirul Islam
1014052074

Contents

Candidate’s Declaration	i
Board of Examiners	ii
Acknowledgement	iii
List of Figures	vii
List of Tables	x
Abstract	xi
1 Introduction	1
1.1 Modeling diabetes	1
1.1.1 Diabetes and its classification	1
1.1.2 Diabetic data collection	2
1.1.3 Types of diabetic model	3
1.2 Research problem	4
1.3 Research gap	5
1.4 Integrated physiological model	5
1.5 Thesis objectives and contribution	7
1.6 Thesis organization	9
2 Literature Studies	11
2.1 Machine learning-based predictive models	11
2.1.1 Type and number of input parameters	11
2.1.2 Classes of used machine learning techniques	12
2.1.3 Performance metrics for model assessment	12
2.1.4 Which ML model is best	15
2.2 Limitations of ML models and way of overcome	15
2.2.1 Limitations of ML models	16
2.2.2 Literature for solutions	17
2.3 Mathematical models/simulators of glucose dynamics	19

2.4	Mathematical models of glucose-exercise dynamics	22
2.5	Research goal	25
3	Methodology	26
3.1	Datasets for modeling diabetes	26
3.1.1	Oral glucose tolerance test (OGTT) data	26
3.1.2	Data on exercise in a clinical experiment	27
3.1.3	CGM data of type-1 diabetic	27
3.1.4	CGM data of type-2 diabetic	28
3.2	ML-based predictive model	31
3.3	Physiology of blood glucose regulation	32
3.3.1	Fasting glucose homeostasis	34
3.3.2	Blood glucose regulation after carbohydrate ingestion	35
3.3.3	Blood glucose regulation after insulin injection	36
3.3.4	Blood glucose regulation at physical activity	37
3.3.5	Rate-balance-concentration paradigm	39
3.3.6	Defining dynamics mathematically	40
4	Integrated Physiological Model	43
4.1	Constraint-based glucose regulation	43
4.1.1	Glucose dynamics	43
4.1.2	Exercise dynamics	44
4.1.3	Free Fatty Acid (FFA) dynamics	44
4.1.4	Lactate dynamics	45
4.1.5	Insulin dynamics	45
4.1.6	Glucagon dynamics	45
4.1.7	Modeling external stimulus	45
4.2	Parameter estimation by model optimization	46
4.3	Reshaping physiological model for OR	48
4.3.1	Glucose dynamics	48
4.3.2	Insulin dynamics	49
4.3.3	Glucagon dynamics	49
4.3.4	Estimating parameters and constraints range	49
5	Experimental Results and Discussion	51
5.1	Fitting OGTT experiment	51
5.2	Fitting clinical exercise experiment	52
5.3	Fitting continuous glucose profile	55
5.4	Comparison with the S.M.Ewings model	57

5.5	Comparison of fitting experiments	62
5.6	Effect of correlation variation in ML models	63
5.7	OR on ML-based forecasting	64
5.8	Discussion	67
5.8.1	Building and validation of the physiological model	68
5.8.2	OR on ML for physiological interpretation	70
6	Conclusion	72
6.1	Summary and findings	72
6.2	Recommendation for future works	73
	References	74
A	Physiological Models of External Stimuli	78
A.1	Gastro intestinal track compartment	78
A.2	Subcutaneous compartment	79
A.3	Lungs compartment	80
B	Real and Synthetic Dataset	81
B.1	Algorithm for synthetic activity profile	82
C	Simulink Implementation of IPM	85
D	Thesis Reproduction	92

List of Figures

1.1	Two strategies of modeling diabetic behavior.	3
1.2	Concept of integrated physiological model.	6
1.3	Simulink implementation of the proposed integrated physiological model (IPM).	7
1.4	Hybrid predictive model of blood glucose dynamics.	8
1.5	Procedure of conducting OR on ML-based glucose forecasting.	9
2.1	The type and number of input parameters used to train the models [3].	12
2.2	Classes of machine learning techniques used in the modeling of blood glucose prediction [3].	13
2.3	The performance metrics used to assess the predictive power of the developed models [3].	13
2.4	Scatter plot of Clarke Error Grid [19].	14
2.5	Composition of operations research.	18
2.6	Sequential steps of OR approach [24].	19
2.7	The user interface of the AIDA diabetic simulator.	20
2.8	Insulin Subsystem of S.M. Ewings model.	23
2.9	Insulin-dependent glucose subsystem of S.M. Ewings model.	23
2.10	Insulin-independent glucose subsystem of S.M. Ewings model.	24
3.1	Dataset of a type-1 diabetic for 04 days: (a) CGM (b) Carbohydrate (c) Activity (d) Insulin.	28
3.2	Daily pattern of glucose variation of a type-2 diabetic patient over CGM of 20 days.	28
3.3	Daily pattern of carbohydrate(mg) taking in logged events over 20 days.	29
3.4	Continuous carbohydrate onboard (COB) profile of 20 days.	29
3.5	Correlation analysis of glucose appearance rate from the gut with glucose profile.	30
3.6	Physical activity profile of a type-2 diabetic patient over 20 days.	30
3.7	Correlation analysis of physical activity signal with glucose profile.	30
3.8	Synthetic activity profile based on CGM of a type-2 diabetic.	31
3.9	Correlation analysis of synthetic activity signal with glucose profile.	31
3.10	Architecture of FFNN based predictive model.	32
3.11	Organs/tissues involved in blood glucose regulation.	33

3.12	Body response during fasting.	35
3.13	Body response after diet ingestion.	36
3.14	Body response after insulin injection.	37
3.15	Body response during the onset of exercise.	38
3.16	Body response just after exercise.	39
3.17	Rate-balance-concentration paradigm in blood glucose regulation.	40
3.18	Hyperbolic tangent function for effect representation.	41
5.1	Trajectory of glucose appearance rate (mg/min) during OGTT.	51
5.2	Real (dashed) vs. simulated (continuous) glucose concentration for OGTT.	52
5.3	Real (dashed) vs. simulated (continuous) insulin concentration for OGTT.	52
5.4	Trajectory of oxygen consumption rate during a clinical experiment of exercise.	53
5.5	Real (dashed) vs. simulated (continuous) plasma glucose concentration for clinical exercise experiment.	53
5.6	Real (dashed) vs. simulated (continuous) plasma FFA concentration for clinical exercise experiment.	54
5.7	Real (dashed) vs. simulated (continuous) plasma lactate concentration for clinical exercise experiment.	54
5.8	Real (dashed) vs. simulated (continuous) plasma insulin concentration for clinical exercise experiment.	55
5.9	Real (dashed) vs. simulated (continuous) glucagon concentration for clinical exercise experiment.	55
5.10	Continuous glucose appearance rate (mg/min).	56
5.11	Continuous activity signal in % of maximum volume of oxygen consumption.	56
5.12	Real continuous glucose profile (dashed) vs. simulated (continuous) glucose profile for type 2 diabetes.	57
5.13	Simulink implementation of S.M.Ewings model.	58
5.14	Single-day protocol of a type 1 diabetic in continuous form: (a) Carbohydrate (b) Insulin (c) Activity.	58
5.15	Comparison among real glucose and simulated glucose signals of a type 1 diabetic for a single day longer.	59
5.16	Liver compartment - Effect of glucagon (a), glucose (b) and insulin (c) on hepatic release. Effect of glucose (d) and insulin (e) on hepatic uptake.	60
5.17	Muscle compartment - Effect of activity (a), glucose (b) and insulin (c) on muscle uptake of glucose.	60
5.18	Pancreas compartment - Effect of glucose CHANGE (a) and glucose (b) on insulin release. Effect of glucose (c) on glucagon release	61

5.19	Plasma compartment - Effect of activity (a) on insulin degradation. Effect of glucose (b) on RBC uptake. Effect of glucose (c) on glucose uptake in nervous system	61
5.20	Producing OR response by optimizing physiological model.	64
5.21	Liver compartment - Effect of glucagon (a), glucose (b) and insulin (c) on hepatic release. Effect of glucose (d) and insulin (e) on hepatic uptake.	65
5.22	Muscle compartment - Effect of activity (a), glucose (b) and insulin (c) on muscle uptake of glucose.	65
5.23	Pancreas compartment - Effect of glucose CHANGE (a) and glucose (b) on insulin release. Effect of glucose (c) on glucagon release	66
5.24	Plasma compartment - Effect of activity (a) on insulin degradation. Effect of glucose (b) on RBC uptake. Effect of glucose (c) on glucose uptake in nervous system	66
5.25	Clarke Error grid analysis of OR Response (Red) and ML Response (Green) for five (05) days.	67
5.26	Peripheral glucose uptake rate during the FSIGT test: type II diabetic patients (-) and healthy subjects (-).	69
5.27	Insulin multipliers in peripheral glucose uptake rate versus normalized plasma insulin concentrations, type II diabetic patients (-) and healthy subjects (-). . . .	69
5.28	Glucose multipliers in peripheral glucose uptake rate versus normalized plasma glucose concentrations, type II diabetic patients (-) and healthy subjects (-). . .	70
A.1	Gastric emptying and appearance of glucose fluxes in circulation.	78
B.1	Hyperbolic tangential relation between exercise and glucose excursion.	84
C.1	Simulink implementation of liver compartment.	86
C.2	Simulink implementation of muscle compartment.	87
C.3	Simulink implementation of adipose compartment.	88
C.4	Simulink implementation of nervous system compartment.	89
C.5	Simulink implementation of pancreas compartment.	90
C.6	Simulink implementation of plasma compartment.	91

List of Tables

4.1	Tabular representation of metabolic relation among plasma variables and metabolic processes.	48
5.1	List of parameters and their estimated values of S.M. Ewings model for single-day glucose profile of type 1 diabetic.	62
5.2	Correlation coefficient among simulated responses generated from proposed model and corresponding target in different fitting experiments.	62
5.3	Accuracy comparison of two FFNN of same architecture trained on two different datasets (real & synthetic) of different correlation over various PH for 05 full Days.	63
5.4	Comparison among ML response vs. target & OR response vs. target over 03 hours PH for 05 full days.	67
B.1	OGTT dataset for proposed model validation.	81
B.2	Age, height, weight, maximum oxygen uptake and work load during prolonged exercise.	82
B.3	Mean values of arterial substrate and hormone concentrations during and after prolonged exercise.	82
B.4	Assumed dataset for establishing relation between exercise and glucose variation.	83

Abstract

Diabetes management consists of two major tasks: forecasting the blood glucose trend and taking a therapeutic decision. Due to the advancement of sensor technologies, it becomes easy to obtain continuous glucose monitoring (CGM) and physical activity data along with logging of diet and injected insulin information. These data of the diabetic patient are being leveraged by applying machine learning (ML) strategies to obtain future trajectories of glucose level which contains no explanation.

This thesis is aimed to produce physiological explanation from ML-based forecasting of glucose concentration by operation research (OR). For producing forecasting, ML-based model is trained using CGM profile with diet and activity information of a type-2 diabetic patient. Due to unavailability in the literature, a constraint-based comprehensive glucose dynamics model integrated with other physiological models of external stimuli is also aimed to build for OR. An integrated physiological model consisting of glucose regulation and models of external stimuli is considered as a composition of several compartments separately connected with a common compartment named ‘plasma’. Plasma is the only accessible compartment and contains the state variables. Plasma variables are the integrated result of the net change in rates of metabolic processes and basal rates are influenced between two saturation constraints for an operating range of each variable. The influence of a plasma variable on a metabolic rate is represented using a form of the hyperbolic tangent function. Validation is done by fitting the model with clinical experiments and CGM data of a free-living environment. A feed-forward neural network (FFNN) being trained on CGM data along with diet and activity log is used to produce forecasting. The proposed constraint-based glucose regulation model of this thesis is optimized on the forecasting of FFNN with sequential quadratic programming using preestimated personalized constraints. The proposed integrated physiological model generates an average correlation coefficient of 0.84 ± 0.12 on all simulated responses with the target in the fitting experiments. Besides this, the model can produce a spectrum of metabolic effects of plasma variables for showing more insight into glucose metabolism. Both OR response and ML forecasting are compared with real glucose profiles. Though increased RMSE is obtained for OR response in comparison to ML forecasting, an acceptable accuracy is found in Clarke Error Grid Analysis. ML-based forecasting of glucose profile is transformed into optimized glucose trend with physiological interpretation. The interpretation is visualized in a metabolic spectrum derived from a constraint-based comprehensive glucose regulation model. The adopted hybrid approach is capable of encapsulating both generalization of ML and the explanation of the physiological approach.

Chapter 1

Introduction

Diabetes is a disease of the modern lifestyle in the era of industrialization where people are less utilizing their bones in their daily life activities. According to the latest data published in the International Diabetes Federation (IDF) Diabetes Atlas 9th edition [1], there are 463 million adults are currently living with diabetes. It is estimated by IDF that 578 million people will have diabetes by 2030 if sufficient preventive actions are not taken and the number will reach 700 million by 2045. IDF has also estimated the annual global health expenditure on diabetes at USD 760 billion. It is projected that these direct costs will reach USD 825 billion by 2030 and USD 845 billion by 2045. It is also estimated by IDF that premature death, disability and other health complications due to diabetes produce an additional 35% of indirect costs to the annual global health expenditures associated with the condition. There are also some intangible costs of diabetes that include worry, anxiety, discomfort, pain, loss of independence, and a host of other crucially important features of living with diabetes. Hence it is a great scope for healthtec researchers for inventing a strategy to minimize cost by leveraging information technology.

1.1 Modeling diabetes

According to the title, this thesis belongs to the scope of quantitative analytics of diabetes diseases. Building a model using the correlation among quantitative information of diabetes and domain knowledge can be very helpful and effective in long-term diabetes management. Hence it is necessary to get a brief introduction on diabetes and its modeling approaches at the beginning of the research.

1.1.1 Diabetes and its classification

In a healthy person, the blood glucose level is regulated between 70-110 mg/dl by two peptide hormones, called insulin and glucagon, secreted from the pancreas. In people with diabetes, this

regulatory mechanism is disturbed due to impairment of insulin secretion or sensitivity. The American Diabetes Association (ADA) defined Diabetes as: “*a group of metabolic diseases characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both*” [2]. As a consequence, the entrance of glucose into muscles and other tissues is reduced and the glucose absorbed from the gastrointestinal track remains in the blood. There are three principal types of diabetes:

1. Type-1 diabetes mellitus
2. Type-2 diabetes mellitus
3. Gestational diabetes mellitus

Type-1 diabetes is an autoimmune disease in which the insulin-producing β -cells of the pancreatic islets are mistakenly attacked and destroyed by the body’s immune system. As a result, little or no insulin is produced. Type-2 diabetes is called non-insulin-dependent diabetes mellitus (NIDDM). This type leads to the permanent prevalence of higher blood glucose because of the presence of insulin resistance and relative insulin deficiency. In gestational diabetes mellitus, increased blood glucose level due to various hormonal effects during the pregnancy is observed. Among different types of diabetes, type 2 diabetes is the most common type affecting 90 to 95% of the diabetes population around the world [2]. Since no permanent treatment of diabetes is invented till today, this disease is needed to be managed to lead a healthy life by consuming the right amount of carbohydrate and regular physical activity in free-living condition. But **diabetes management** in a free-living environment is a challenging task compared to a clinical condition. Measuring blood sugar level and taking an appropriate amount of food\exercise\medication based on predicted future trends transform diabetes management as prediction (forecasting) and decision (optimization) generating process technically.

1.1.2 Diabetic data collection

Usually, the glucose level is estimated with a device taking a single drop of blood by a finger prick. This invasive technique can’t be conducted frequently and give a small amount of data regarding blood glucose variation from a diabetic patient. Due to the invention and usage of Continuous Glucose Monitoring (CGM) sensors, which measure subcutaneous glucose concentration from the interstitial fluid at a regular interval (1-5 min), a large volume of information regarding blood glucose readings and their trend can be available in free-living condition. At the same time tracing physical activity level and energy expenditure has become easier due to the availability of smart-watches, fitness bands, and other body area networks. Besides these, logging down the ingredients of each diet and injected insulin regimen is possible leveraging the smart-phone application.

1.1.3 Types of diabetic model

An attempt of building a personalized diabetic model using diabetic data obtained through sensor technologies and logging has already been done. That model can be very useful in diabetes management for both physicians and patients in approximating food taking, exercise duration and intensity, and a dose of insulin therapy. A diabetic model can also be used in the controller module of model predictive control in Artificial Pancreas System (APS) development.

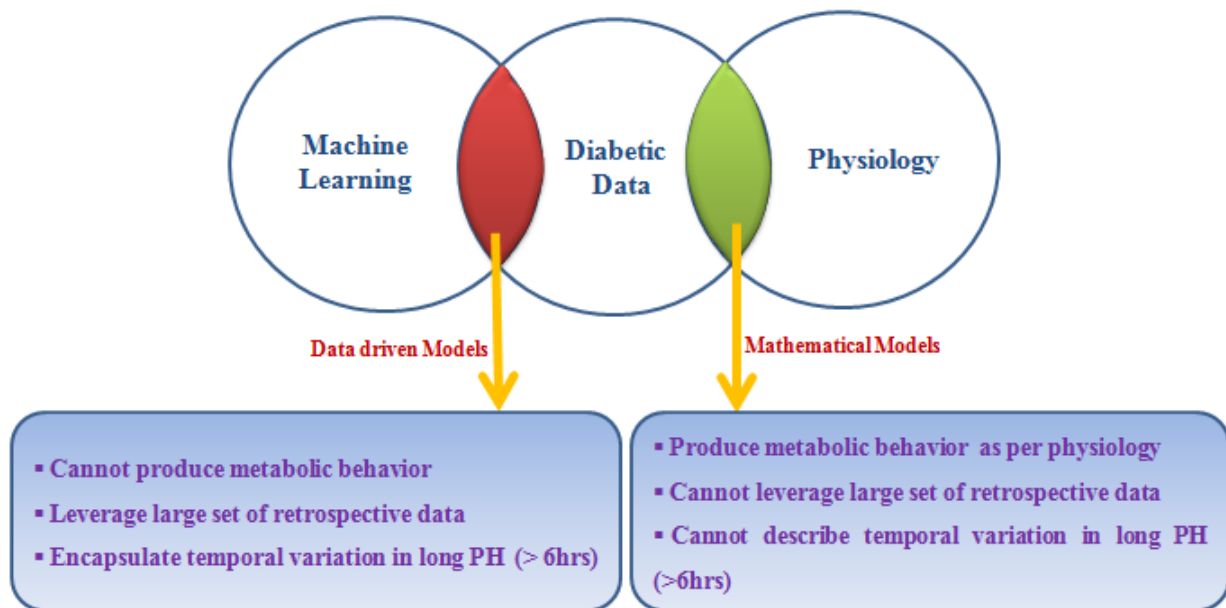


Figure 1.1: Two strategies of modeling diabetic behavior.

Usually, two basic types of strategy are appeared in the research field of diabetes modeling (Figure 1.1). In a **data-driven approach**, diabetic data consisting of continuous blood glucose measurements obtained from CGM sensors and logged life style information are used to build a predictive diabetic model. By using Machine Learning (ML) upon retrospective observations, a hidden pattern of blood glucose variation is exploited and glucose concentration is predicted without any knowledge of the underlying physiological processes. The ML-based predictive model relies on some non-physiological formulations to characterize the relationship between current and past CGM values. These models usually consider the task of predicting blood glucose as a time series forecasting problem since blood glucose measurements have a natural temporal ordering. The main drawbacks of the ML-based model are the absence of causal reasoning, poor performance on the infrequent glycemic excursion, and lack of physiological interpretation [3], [21], [36].

In the **physiological modeling approach**, blood glucose dynamics are represented by mathematical equations based on prior physiological knowledge of metabolic behavior. Parameters of the model are estimated by optimization on data from a clinical experiment

on a group of persons. These models usually show poor performance in producing time-variant metabolic behavior over the long term (> 6 hours) [36].

1.2 Research problem

A study on the literature of ML-based forecasting models [3] of blood glucose says that almost all types of ML techniques such as neural networks, support vector machines, decision trees, and other time-series forecasting methods have already been applied to build the predictive model. But, to the best of the author's knowledge still, these models are not seen to be applied in making any therapeutic decision for diabetes management. The major limitations are the absence of robustness analysis against noisy inputs and physiological interpretability in the forecasted glucose trajectory. ML models leverage the correlations among the sequences of attributes of a dataset to estimate the glucose concentration at future time points. If correlations in the dataset vary due to measurement noise or missing records, the performance of ML models is also affected. Besides that, the ML models usually work as the black-box classifiers. When a forecast is obtained over a prediction horizon (PH), there is no physiological idea of how the result is being generated. That's why ML-based models are not found to be useful for both patients and health professionals.

For robustness analysis of ML-based models against noisy inputs, synthetic dataset with noise and discontinuous sequences can be adopted. But, *what type of physiological interpretation is required to make a therapeutic decision?* In searching for the answer to this question, models used to make decisions in diabetes management are studied. It is found in the literature [4]-[7] that most of the predictive models of glucose dynamics used to estimate insulin dose are mathematical. The parameters of those models represent specific metabolic characteristics providing physiological realization. But, on the other hand, those mathematical models can't take advantage of hidden patterns laid in a large amount of retrospective CGM data similar to ML models. To the best of the author's knowledge, no attempt is seen to produce physiological interpretation from ML-based forecasting.

Based on the above description, a solution is imagined for removing the limitation of the ML-based predictive model to facilitate diabetes management. If a segment of significant length (≥ 03 hours) of a forecasted profile of an ML-based predictive model can be fit in a mathematical structure with personalized constraints through mathematical programming, it is possible to regenerate that profile segment with causal reasoning and physiological interpretation. This process is known as operation research (OR) and the application of OR in the research field of diabetes is also available in the literature [8]. The nature of anticipated physiological interpretation completely depends on the architecture of the adopted mathematical model.

1.3 Research gap

In this thesis, a dataset of a type-2 diabetic patient consisting of CGM profile, diet events, and physical activity log is used to build an ML-based predictive model for conducting OR described in the previous section. Hence, a constraint-based physiological model of glucose dynamics for the type-2 diabetic considering impact of exercise is required for applying in OR. At the commencement of this research based on a rigorous study, it is found that existing models [13]- [16] of diabetes with exercise dynamics are built on the assumption of no endogenous production of insulin. Usually, type-1 diabetic patients cannot produce endogenous insulin and are completely dependent on external insulin injections. So, those models are not capable of describing exercise dynamics for insulin-independent diabetic patients or type-2 diabetic and a research gap is obvious.

To fill up the gap a constraint-based physiological model of glucose regulation consisting of internal insulin dynamics and external stimulus of diet, insulin, and exercise is needed to be built. This sort of model is significant for two (02) more reason as below. As stated earlier, more than 90% of diabetic are of type-2 and manage their glucose level in free-living condition. A type-2 diabetic person may have impaired insulin secretion from the pancreas or insulin resistance in hepatic/peripheral tissues [8]. As a result, glucose concentration persists at a level higher than normal for a longer period which is called hyperglycemia. Physical activity has a great short and long-term impact on blood glucose regulation. Muscle tissue uptakes glucose from circulation without the influence of insulin during the onset of exercise and helps to control blood glucose levels in the normal range [13]. Hence, regular physical activity is prescribed for diabetes management in free-living conditions. To cover the scope of free-living environment for the type-2 diabetic population the effect of physical exercise on glucose dynamics is required to incorporate.

Another important issue is to perform robustness analysis of ML-based predictive model inside an *in silico* environment against noisy and missing data. Multiple sets of synthetic data with variety of correlation and discontinuous sequences are required to serve the purpose. For preparing in *in silico* environment a diabetic simulator which represents physiology of blood glucose regulation consisting of internal insulin secretion and external stimulus of diet, insulin, and exercise is needed.

1.4 Integrated physiological model

From the previous section it is learned that a physiological model of blood glucose regulation with some specific features is required to solve a particular problem of ML-based approach in modeling diabetes. This type of model is also significant for type-2 diabetic because of being large in number than other types of diabetes. At the same time, desired model is necessary to

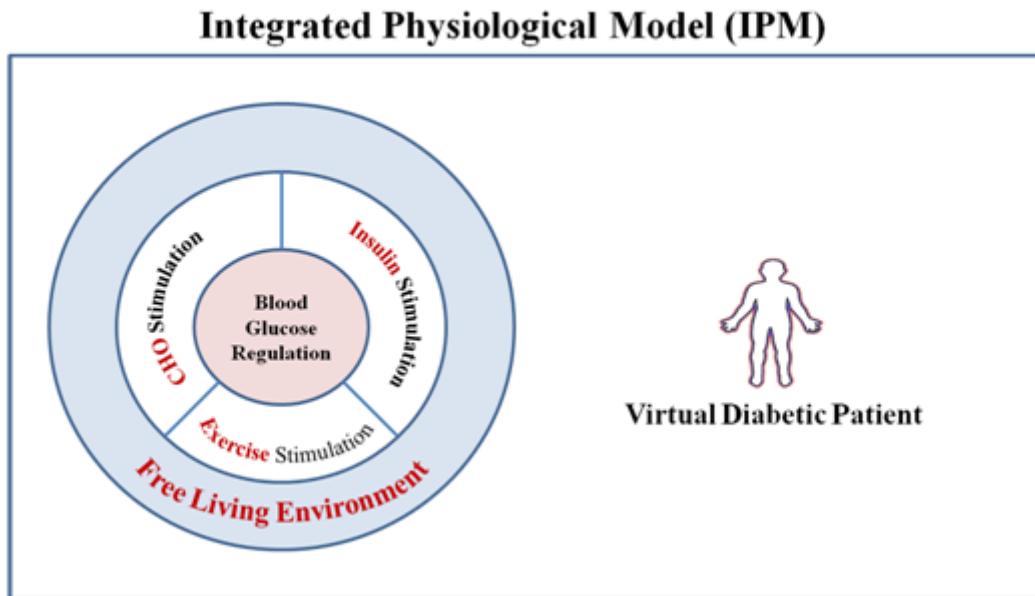


Figure 1.2: Concept of integrated physiological model.

produce synthetic glucose profile for both type-1/type-2 diabetic in free living environment. To serve all these utility a comprehensive, multipurpose, easy to understand, and composite physiological model of glucose dynamics is needed. But this type of composite model should be consisted with glucose dynamics and physiological models of external stimulus for being complete.

In free-living environment glucose dynamics are primarily interrupted with three (03) categories of events: diet, insulin, and physical activity. These events are occurred in irregular periodicity in the daily life of a diabetic patient. The nature and volume of glycemic excursion is directly proportional to the magnitude and type of those events which are defined as external stimulus in the physiology of glucose regulation. So, for free-living environment, external stimuli are the glucose appearance into circulation due to oral ingestion of carbohydrate, insulin appearance into circulation from the site of subcutaneous injection and oxygen consumption due to physical movement. External stimulus does not influence the glucose regulation instantaneously rather need to pass through a physiological process. Physiological models for representing the impact of external stimuli on glucose regulation with the physiological delay are available in the literature [13], [31], [32]. The integration of plasma glucose regulation with those physiological models of external stimulation is required for the structure of a complete organization (Figure 1.2). This organization can be defined as the integrated physiological model (IPM).

When this IPM is implemented in computer environment for simulating diabetic behavior, it mimics as a virtual diabetic patient. Usually, the virtual diabetic patient is the implementation of a system of differential equations and parameters for representing impaired glucose metabolism in the human body [9]- [12]. By tuning the parameters, it is possible to describe the metabolic disorder of diabetic patients with inter and inpatient characteristics in a computer environment.

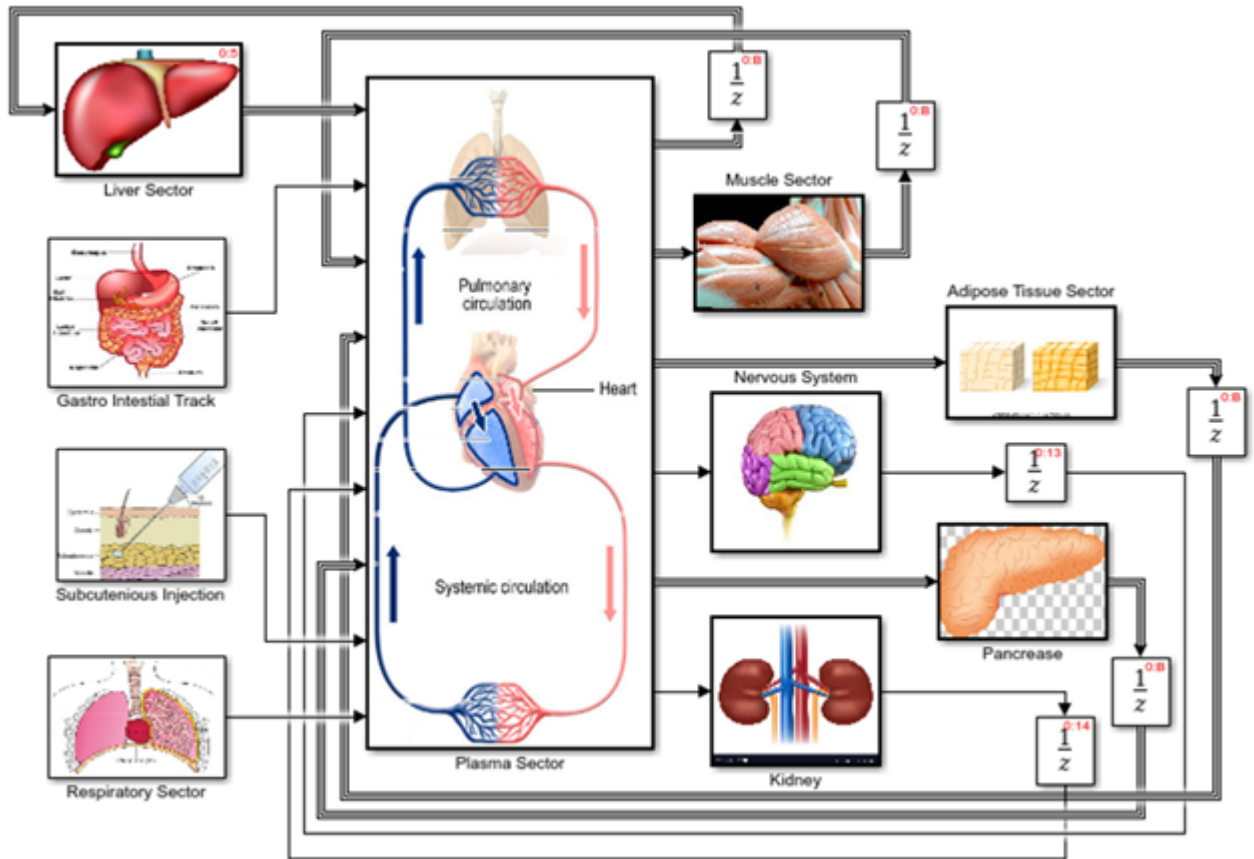


Figure 1.3: Simulink implementation of the proposed integrated physiological model (IPM).

The Figure 1.3 shows the MATLAB/Simulink implementation of the concept of IPM. The gastro intestinal track, subcutaneous injection, and respiratory compartment are implemented using the existing physiological model available in the literature. Rests of the compartments of the IPM are concerned with the proposed constraint-based glucose regulation. The IPM of Figure 1.3 is assumed to be a complete simulator and capable of producing abnormal metabolic behavior of different organs involved in the glucose regulation of a diabetic patient (type-1/type-2) which are not possible in a minimal model.

1.5 Thesis objectives and contribution

The objective of this thesis is the implementation of the imagined solution (section 1.2) which is OR on ML-based forecasting to produce a hybrid model combining the advantages of both data-driven and physiological approaches (Figure 1.4).

A feed-forward neural network (FFNN) based predictive model is built with CGM data along with diet and activity information of a type-2 diabetic patient to produce ML-based forecasting. A constraint-based mathematical model of glucose regulation is adopted for OR. The OR model is optimized with real CGM data along with protocol information to estimate the limit of the

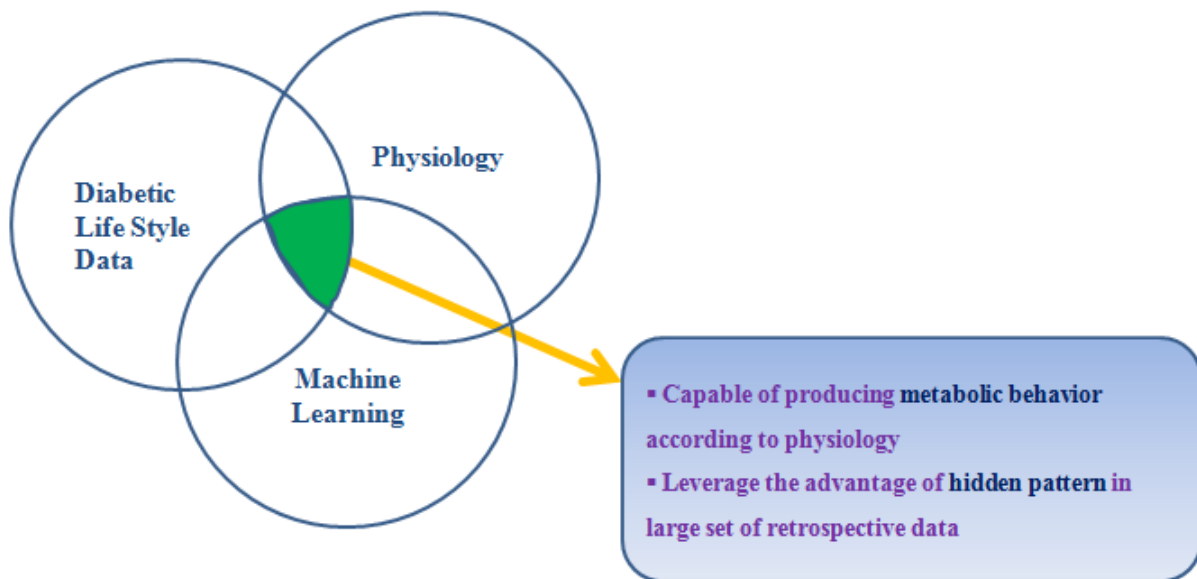


Figure 1.4: Hybrid predictive model of blood glucose dynamics.

constraints. Then, forecasted sequences of FFNN are optimized in the OR model with the estimated limit of the constraints to reproduce forecasted glucose profile with a physiological explanation. Figure 1.5 describes the complete procedure schematically. The fluctuation of both OR and ML responses from real CGM series are estimated in terms of RMSE and placed in Clarke Error Grid analysis. It is observed that RMSE is increased due to optimization through a lot of constraints and consideration of less number of plasma variables. But encouraging results are obtained in error grid analysis since maximum points are located in the clinically acceptable zone.

Due to the gap in the literature, this thesis also aims to build a virtual diabetic patient which is basically an integrated physiological model (IPM). This IPM integrates a constraint-based comprehensive glucose dynamics model with existing physiological models of external stimulus. This IPM is capable of describing desired metabolic behavior of diabetes and also plasma glucose variation over carbohydrate ingestion, insulin injection, and exercise events. The glucose regulation model of IPM is considered to consist of several compartments. The plasma compartment of IPM is assumed to contain the state variables. State variables influence the basal rate of metabolic processes (uptake/release/ secretion/degradation) of all compartments connected with plasma. The influence of state variables on metabolic rates is nonlinear in nature and represented using a form of the hyperbolic tangent function.

The adopted function can represent threshold and saturation phenomena which are typically observed on a metabolic process due to a change in concentration of one or more substrates [17]. Moreover, the influence of the controlling variables (e.g. hormones) can be easily incorporated into these relationships as a change in the saturation value and the slope of the trajectory. The main rationale to apply hyperbolic tangent function is to introduce physiological and saturation

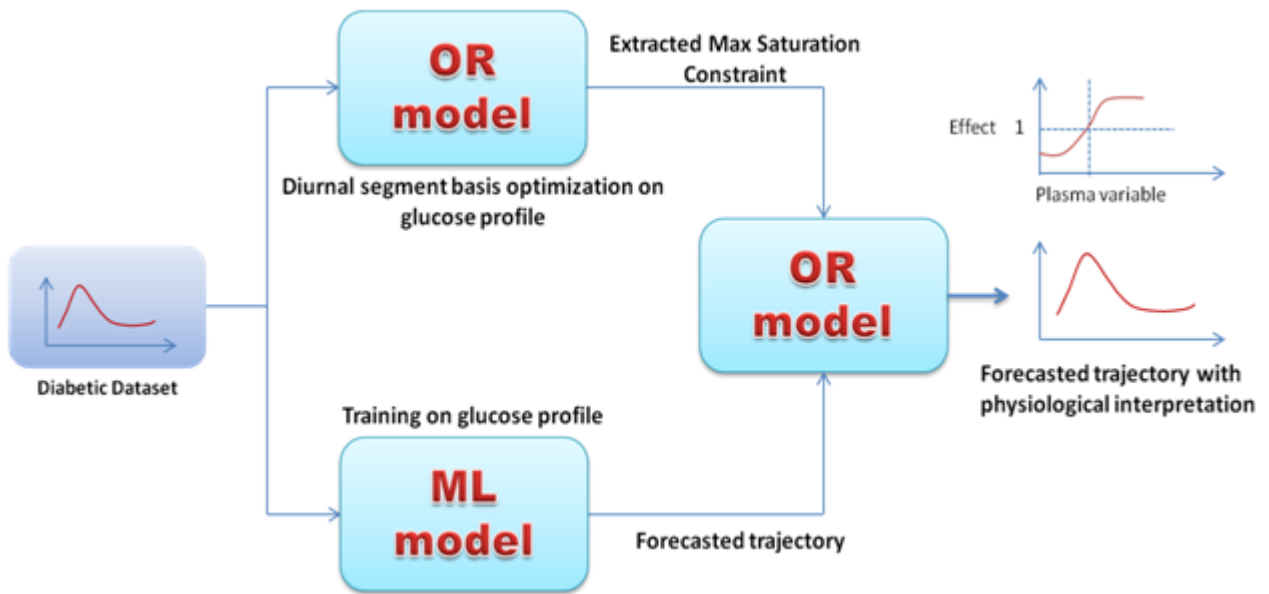


Figure 1.5: Procedure of conducting OR on ML-based glucose forecasting.

constraints. The parameters of the model are tuned through constrained nonlinear optimization and the model can be used in operation research (OR). Implementing this constraining feature using a linear differential equation (DE) is difficult. The Parameters of the DEs are static rate constant and provide less implication. In fitting experiments on various experimental datasets, the proposed model produces an average correlation coefficient of 0.84 ± 0.12 on all simulated responses with the target. Besides this, the optimized model can generate a metabolic spectrum by plotting effect equations for providing more metabolic insight.

1.6 Thesis organization

Chapter 2 shows the justification behind the research problem and research gap mentioned in the introduction by studying the literature. Before going for the justification, chapter presents a bird's eye view of the research field of ML-based glucose forecasting models. This chapter finds out the limitations of ML-based predictive models and shows potential solutions. By describing the merits and limitations of the existing prominent glucose dynamics and exercise models along with their characteristics and scope, this chapter ends by setting the research goal of this thesis.

Chapter 3 describes the datasets of clinical experiments and free-living conditions used for building and validating ML and physiological models. This chapter also illustrates the architecture of FFNN for building a predictive model and physiology of blood glucose regulation during disturbances of carbohydrate ingestion, insulin injection, the onset of physical activity, and fasting event.

Chapter 4 describes the mathematical structure of the proposed integrated physiological model in detail in terms of constants, parameters, state variables, and equations. This chapter also explains the parameter tuning process of the model.

Chapter 5 deals with the result of fitting experiments, comparison with a reference model, and the result of OR on forecasting of FFNN. It also includes the discussion on conducted experiments with the proposed integrated physiological model and the application of the model in OR. The discussion usually explains why and how the result of the experiments is produced. Limitations of the model are also described in the discussion.

Finally, Chapter 6 summarizes the thesis and provides recommendations for future research work.

Chapter 2

Literature Studies

The research problem of this thesis is defined explicitly in the introduction focusing on a research gap in the literature. But it is essential to justify the defined problem and proposed solution by exploring the existing research. This justification is the main objective of this chapter. Before going for the assessment of the limitations of the ML-based predictive models, it is helpful to obtain a high-level view of the concerned field of blood glucose forecasting. Discussion on existing analogous models is very important to build a constraint-based physiological model of glucose regulation consisting of external stimuli.

2.1 Machine learning-based predictive models

The task of presenting an overview of the research field of machine learning-based blood sugar forecasting is become easy by studying the review paper authored by A.Z. Woldaregay et.al. [3]. In this review, a total of 55 articles are selected after a full-text assessment out of 624 papers which are retrieved from rigorous searching between August 2017 and February 2018 in various online databases including Google Scholar, PubMed, ScienceDirect, and others. The selected articles were compared based on some categories which are defined on rigorous brainstorming and discussions. Categories are age and number of subjects, type of input, data format or type/data source, input preprocessing, class of machine learning, training/learning algorithm, validation techniques, prediction horizon (PH), performance metrics.

2.1.1 Type and number of input parameters

According to the reviewed literature, blood glucose, insulin, and diet comprise the most used group of parameters (29%), as shown in Figure 2.1. BG, insulin, diet, and physical activity make up the second most used group of parameters (25%). The use of only the BG parameter ranked third (20%). The use of BG, insulin, diet, and physical activity, stress and, other groups

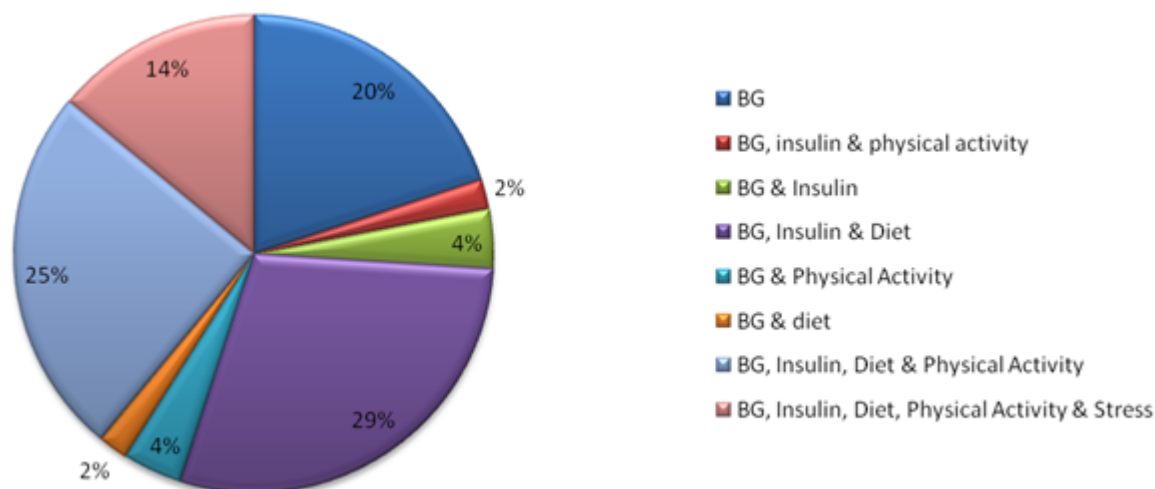


Figure 2.1: The type and number of input parameters used to train the models [3].

of parameters ranked as the fourth most used (14%). The use of BG along with insulin and BG along with physical activity ranked equally as the fifth most used group of parameters (4%). The use of BG with diet and BG with insulin and physical activity ranked equally as the sixth most used parameters (2%).

2.1.2 Classes of used machine learning techniques

Various classes of machine learning techniques have been used in general dynamic system modeling, regression, and prediction services. However, for BG prediction, feed-forward neural networks are the most used techniques (20%), as shown in Figure 2.2. The hybridization of the physiology-based model and machine learning techniques is the second most used approach (19%). Recurrent neural networks in various forms ranked as the third most used technique (18%). Support vector machines (SVMs) ranked as the fourth most used technique (11%). Genetic programming techniques, most notably grammatical evolution, ranked as the fifth most used technique (6%). Autoregressive neural networks and neuro-fuzzy networks are the sixth most used techniques (5%). Self-organizing maps (SOMs) ranked seventh (4%). Extreme learning machines, kernel functions, Gaussian processes, genetic algorithms, and random forests ranked eighth (2%). Jump neural networks and deep neural networks ranked ninth (1%).

2.1.3 Performance metrics for model assessment

Performance metrics are necessary steps that should be carefully chosen based on the developed model under consideration. Various performance metrics are used to assess the predictive power of the developed model. However, choosing the appropriate metrics depends on the type of

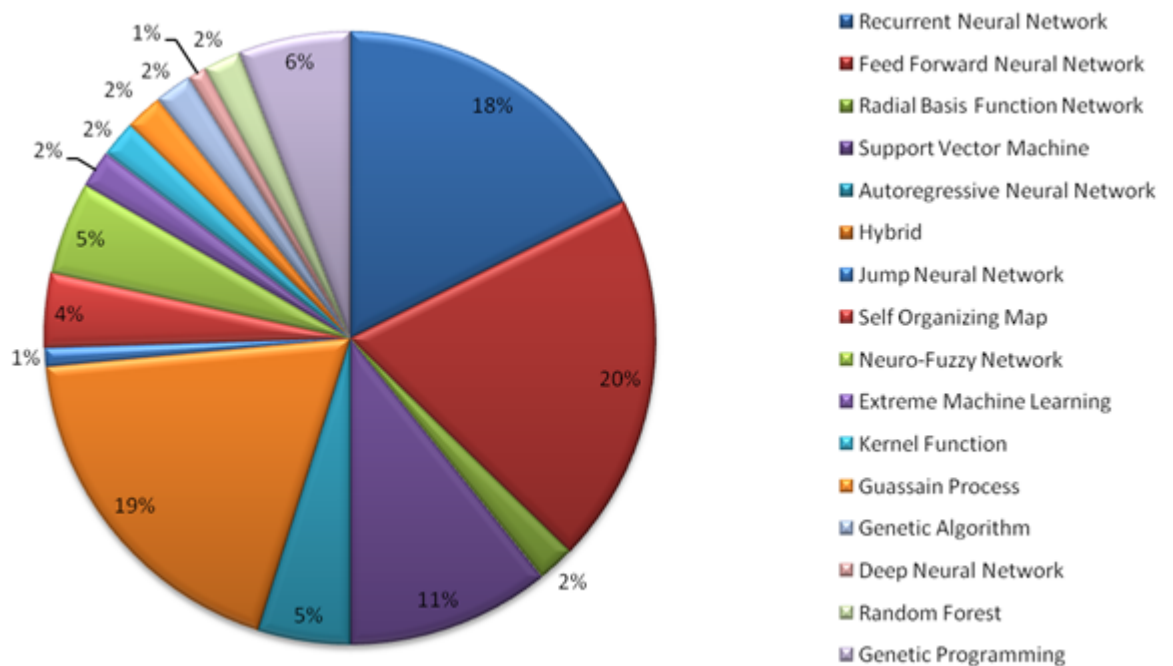


Figure 2.2: Classes of machine learning techniques used in the modeling of blood glucose prediction [3].

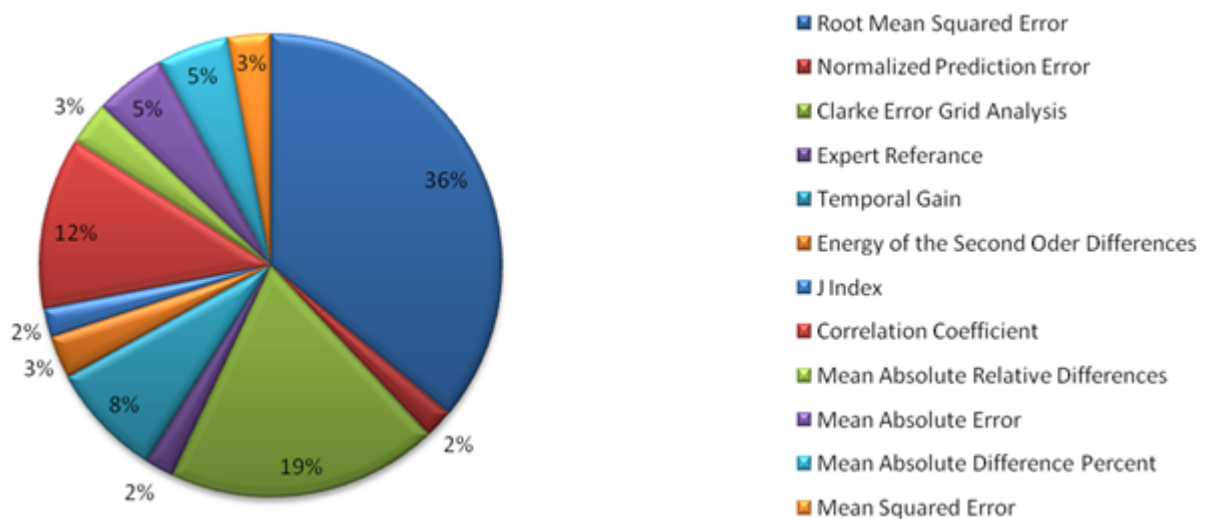


Figure 2.3: The performance metrics used to assess the predictive power of the developed models [3].

application that the model is intended to be used. Based on the reviewed articles, the performance metrics used to assess the predictive performance of the model can be categorized into two groups: mathematical evaluation criteria (empirical accuracy) and clinical evaluation criteria (clinical accuracy). The mathematical evaluation criteria (empirical accuracy) are simply used to evaluate the numerical accuracy without giving due consideration to the clinical significance.

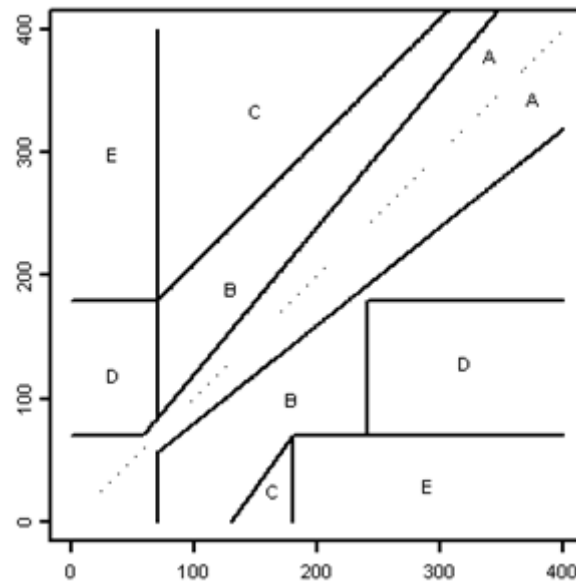


Figure 2.4: Scatter plot of Clarke Error Grid [19].

This group of metrics includes root mean square error, correlation coefficient, FIT, normalized prediction error (NPE), and geometric mean. The clinical evaluation criteria (clinical accuracy) give due consideration to their significance in terms of clinical usability and include error grid analysis, average time gain, mean absolute relative difference, expert comparison, and J index. Generally, the most popular performance metric is the root mean square error (36%), followed by Clarke error grid analysis (19%), as shown in Figure 2.3. The third most popular metric is the correlation coefficient (12%), followed by the temporal gain (8%). The fifth most popular metric is mean absolute error and mean absolute difference percent (5%). The sixth most-used metrics are mean absolute relative difference, the energy of the second-order difference, and mean squared error (3%). The seventh most-used metrics are normalized prediction error, expert reference, and J index (2%).

The Clarke Error Grid Analysis (EGA) was developed in 1987 to quantify the clinical accuracy of patient estimates of their current blood glucose as compared to the blood glucose value obtained in their meter. It was then used to quantify the clinical accuracy of blood glucose estimates generated by meters as compared to a reference value. Eventually, the EGA became accepted as one of the “gold standards” for determining the accuracy of blood glucose forecasting in the ML-based research field of diabetes. The grid breaks down a scatter plot of a reference glucose meter and an evaluated glucose meter into five regions Figure 2.4:

- Region A are those values within 20% of the reference,
- Region B contains points that are outside of 20% but would not lead to inappropriate treatment,

- Region C are those points leading to unnecessary treatment,
- Region D are those points indicating a potentially dangerous failure to detect hypoglycemia or hyperglycemia, and
- Region E are those points that would confuse treatment of hypoglycemia for hyperglycemia and vice versa.

2.1.4 Which ML model is best

Though the review study stated above gives an overview regarding the inputs, ML techniques, and performance metrics, it doesn't give any idea regarding the strength of modeling approaches used so far in the literature. Because different modeling approaches in research works utilize different datasets of different diabetic patients. In one of the recent studies by J. Xie et. al. [20], a comparative analysis has been conducted on various machine learning and classical autoregression techniques based on the OhioT1DM dataset, which includes eight week's data collected from six anonymous T1D patients under insulin pump therapy, where Medtronic 530G insulin pumps with Medtronic Enlite CGM Sensors were used. Besides that life-event data of each subject were reported through a custom smartphone application and activity data were collected from a fitness band. Because of using the same data and feature sets for all the ML techniques, it is possible to assess the strength of all techniques in a fairway.

The major findings from the research are that no significant advantage has been observed from the ML models compared to the classical Auto Regression with Exogenous Inputs (ARX) except that Temporal Convolution Network (TCN) model's performance was more robust for BG trajectories with spurious oscillations. Compare to that, ARX shows over-predict peak BG values and under-predict valley BG values. A major limitation of this attempt is the length of the prediction horizon (PH) which is only thirty (30) minutes. Thirty minutes is not enough time to assess various algorithms of predictive modeling rigorously. Normally, the performance of predictive models deteriorates in the case of longer PH. Hence, if the experiments could be enhanced for more than two hours, it would have been possible to assess the performance more rigorously.

2.2 Limitations of ML models and way of overcome

From the brief review described in the previous section, it is clear that the research field of blood glucose forecasting based on ML has achieved a remarkable base. But still, these models cannot be applied in making the therapeutic decision for diabetes management. By studying the nature of the forecasting models and research works related to estimating the magnitude of action (amount of carbohydrate/insulin) for diabetes management, it is appeared that two logical reasons prevent ML models from being applied in decision making for blood glucose

management. The first reason is the lack of robustness analysis and the second is the absence of physiological interpretability. The only usage of the ML-based predictive model is to get a notification on upcoming hypoglycemic events [35].

2.2.1 Limitations of ML models

The first limitation is the lack of robustness analysis of ML models in presence of noise and missing events. ML strategy has a great generalization capability to extract a pattern from a large number of instances that occurred in the past. Usually, ML-based models leverage the presence of autocorrelations and cross-correlations in retrospective time series data of continuous blood glucose concentration and other factors affecting diabetes. Hence variation of correlations has a severe impact on the performance of the model. The crucial task in diabetes management is the collection of lifestyle data which consists of carbohydrate consumption, physical activity tracking, insulin injection information, mental and physical condition, etc. Usually, these data contain an unbelievable amount of noise variation and discontinuity introduced from various types of sensor technology and manual recording of events. The predictive model that is capable of producing real-time acceptable forecasting must have the mechanism to handle the noise variation and discontinuity events. In chapter 5, an experiment has been conducted on a diabetic dataset (consisting of CGM data, diet, and activity events) to estimate the effect of correlation variation on the performance of a feed-forward neural network (FFNN) based forecasting model. By creating a different synthetic physical activity signal using a sigmoid function of the CGM series (Appendix B), two activity signals of different correlations have been produced. Two FFNN of the same architecture have been trained to build the models and to observe the performance difference due to correlation variation. It is found that increased correlation has better performance in the forecasting of future trajectories of plasma glucose concentration (Table 5.3).

The second limitation is the absence of physiological interpretability in the forecasted glucose trajectory. Parameters of the ML models contain weights that are a bunch of 1's and 0's. These weights do not provide any sense regarding the blood glucose dynamics of the human body. Weights are only interpretable by a machine for producing generalized patterns based on input. Forecasting is mere curve fitting and there is no explanation of how the result is being generated. That's why ML-based models are not found to be used in therapeutic decision making such as estimating the insulin dose required for maintaining a normal glucose range. Due to the absence of causal reasoning ML models usually overestimates hypoglycemic events and underestimates hyperglycemic events in predicted glucose trajectory [21].

Most of the predictive models of glucose dynamics that are used to estimate insulin dose are the physiological model. The parameters of those models have concrete physiological meaning and can be measured for a certain patient in complex clinical protocols. In P. Gyuk

et. al. [4] a prediction algorithm has been proposed for outpatients without taking physical activity under consideration. The model consists of two state-of-the-art physiological models that calculate nutrition absorption and plasma glucose control with insulin evolution. A genetic algorithm has been used to estimate the parameter set. In C.Liu et. al. [5] a glucose forecasting algorithm, based on a compartmental composite model of glucose-insulin dynamics, has been proposed which is currently being used as the core component of a modular safety system for insulin dose recommender developed within the EU-funded project. Matlab *fmincon* constrained optimization routine was employed to estimate the state of the model from the continuous glucose monitoring (CGM) signal. In A. N. Sveshnikova et. al. [6], a mathematical model has been constructed with six ordinary differential equations that describe the dynamics of changes of glucose concentration as well as insulin and anti-insulin factors considering the main physiological parameters of blood-glucose regulation. The parameters have been identified according to continuous glucose-monitoring data using an evolutionary programming method. In Yan Zhang et. al. [7], a data-driven nonlinear stochastic model based on second order differential equations has been developed to describe the response of blood glucose concentration on food intake using CGM data. A Bayesian learning scheme was applied to define the number and values of the system's parameters by iterative optimization of free energy. The four mathematical models stated above are constructed based on the physiology of blood glucose regulation and capable of producing glucose forecasting leveraging a short amount of CGM data with the physiological interpretation. Though these models can't take advantage of the hidden pattern of the large amount of retrospective CGM data, these are more acceptable because of their illustration. Hence, to make the ML-based forecasting effective in diabetes management causal reasoning is required to be produced.

2.2.2 Literature for solutions

In this thesis, a solution has been pursued to overcome those two barriers mentioned in the previous subsection so that ML-based models can be applied in producing therapeutic decisions for diabetes management. For the first case, the performance of ML models can be tested on in silico data with various possible correlations and can be rebuilt to produce an acceptable result. The diabetic simulators/glucose dynamics models can play an important role in this regard by generating a synthetic dataset from a preconfigured virtual diabetic patient. A diabetic simulator produces glucose variation signals based on a prescheduled protocol for a particular set of physiological parameters that represent a virtual patient. By adding noise signals of different magnitude with the simulated glucose concentration and choosing various physiological delays for the parameters, it is possible to generate different sets of data representing correlation variation.

In a recent predictive model by J.Martinsson et.al. [22], noise experiments have been conducted

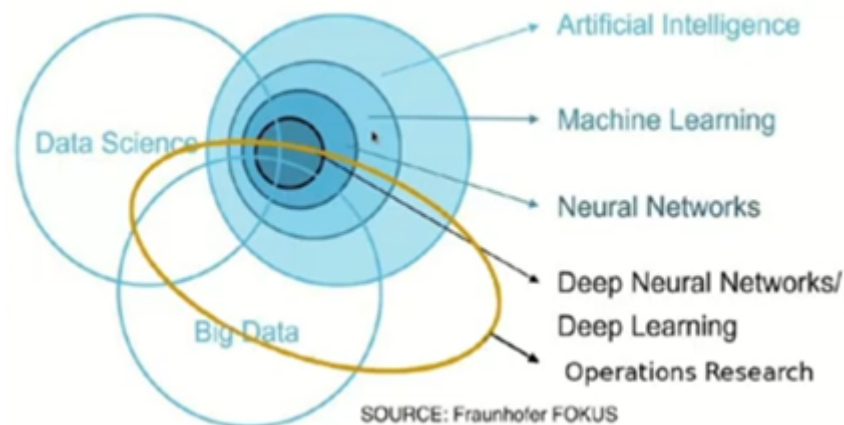


Figure 2.5: Composition of operations research.

to observe what uncertainty the model can learn. Two types of noise called measurement noise and state length noise, have been added to the deterministic signal obtained from the OhioT1D dataset to produce a noisy dataset. OhioT1D dataset isn't freely available. Data Use Agreement with Ohio University at the institutional level is required to obtain access to the dataset. Hence, in absence of real diabetic data, a simulator might be a better alternative for producing synthetic noisy datasets.

For the second case, to provide physiological interpretability of the glucose concentration forecasted from the ML-based predictive model, Operation Research (OR) might be an innovative solution (Figure 2.5). The ML-based forecasted series is the generalized statistical response of glucose concentration. By conducting OR with a constraint-based physiological model upon the response of the ML model, it is possible to transform the statistical signals into physiological/metabolic response.

A review paper by Y. Bengio et. al [23] shows various strategies of hybridization on ML and OR to solve complex problems. The OR approach to solve a problem comprises six sequential steps (Figure 2.6). It is obvious from the Figure 2.6 that to perform operation research on glucose profile predicted from the ML approach, a well-descriptive mathematical model of blood glucose regulation in a healthy human body is required.

Leveraging OR it is not only possible to transform the statistical signal into a metabolic signal but also possible to get an idea regarding the metabolic disorder that occurred in the internal organs of the body. A Ph.D. thesis by O.Vahidi [8] regarding the assessment of organ dysfunction in a type-2 diabetic is available in the literature. That thesis describes how the origin of metabolic disorder in a type-2 diabetic patient can be assessed by performing OR on a clinical experimental dataset with a physiological model of blood glucose regulation.

So, from the above discussion, it is obvious that the two problems have a common solution. The solution is applying a mathematical/physiological model of blood glucose regulation in two

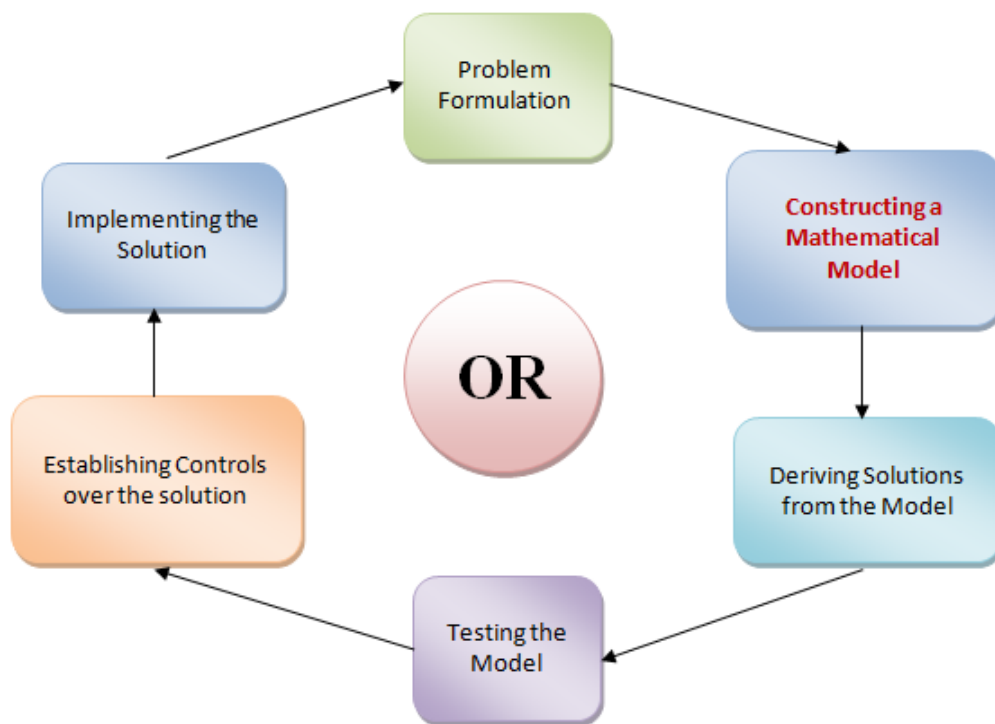


Figure 2.6: Sequential steps of OR approach [24].

different ways. In the first case, the model should be used to produce a noisy synthetic dataset and in the second case, the model shall be used to perform OR.

2.3 Mathematical models/simulators of glucose dynamics

Since mathematical model is assumed as a tactical solution for the two limitations of the ML-based predictive models, it becomes essential to get an overview of recognized mathematical/physiological models available in the literature to obtain an idea regarding the scope, strength, and limitations of these models. [AIDA](#) is a freely available diabetic simulator by developed and maintained by E.D. Lehmann and T.Deutsch. It is suitable for the patient and medical staff education for glucose-insulin interaction in insulin-dependent (type 1) diabetes mellitus.

Over 0.638 million diabetes simulations have been run at AIDA online, from 115 countries all over the world up to January 2014. Neither physical activity nor type-2 diabetes is modeled in the simulator. A snap of the user interface of the AIDA simulator is shown in [Figure 2.7](#).

Among other simulators, the Universities of Virginia and Padova research group model, named as **UVA/Padova model** [9] consists of two subsystems, namely the glucose and insulin system. In a later publication, an additional glucagon subsystem is presented. The glucose subsystem is divided into three systems: the transport, the production, and the utilization of glucose. This model includes the subcutaneous insulin kinetics to simulate the administered insulin. The

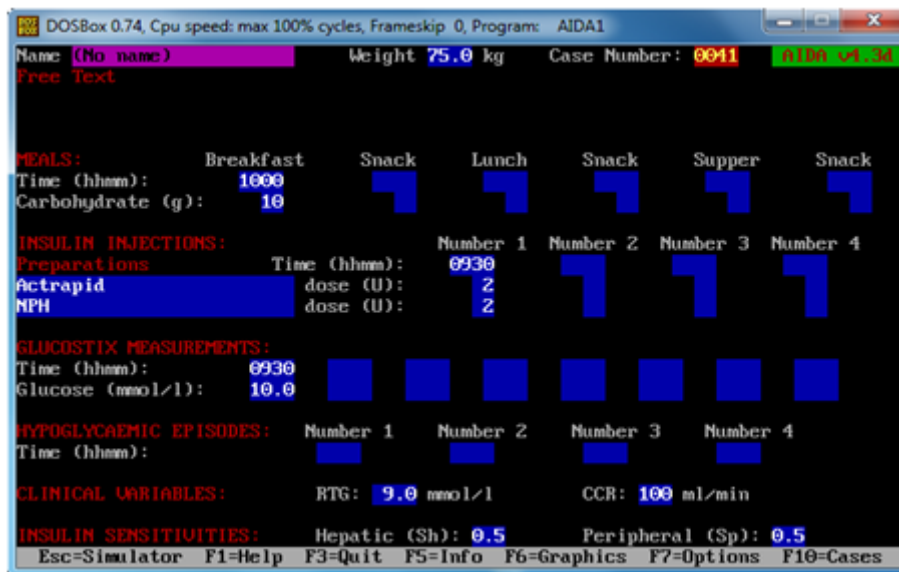


Figure 2.7: The user interface of the AIDA diabetic simulator.

simulation population consists of 300 virtual patients including adults, adolescents, and children. The model is approved by the Food and Drug Administration to replace animal trials.

The **Cambridge's model** [10] consists of five sub-models: the glucose and insulin kinetics, the glucose absorption, the subcutaneous insulin, and the interstitial glucose. With these extended sub-models, the model is specifically built to support the development of a closed-loop system. Bayesian parameter estimation process has been applied to determine time-varying model parameters. The model population consists of 18 virtual patients. The model is validated with an overnight clinical study.

The **Sorensen model** [11] is an explanatory physiological model of glucose metabolism. The model represents the organs in six compartments. These compartments are again divided into three spaces: the capillaries, the interstitial, and the intracellular space. In these spaces, the interactions of glucose, insulin, and glucagon are described. These are represented as a mass balance. This model was the first complete model that simulates an average patient with type 1 diabetes. However, it simulates intravenous administered insulin. Therefore, the delay when insulin is infused subcutaneously is neglected.

The **Fabietti model** [12] is developed by a research group at the University of Perugia in Italy based on a modified Bergman's minimal model [25]. In this model, the endogenous insulin secretion is substituted by subcutaneously delivered exogenous insulin and the glucose kinetics is represented by two instead of a single compartment. External inputs of the model such as meals and intravenous glucose boluses have been added together with the sub-model of the glucose absorption from the gastrointestinal tract. An interesting feature of the model is the sinusoidal representation of the circadian variability of insulin sensitivity. The amplitude and phase of the circadian rhythm are estimated 'off-line' to characterize an individual subject.

Comparing all the models stated above, the Cambridge and UVA/Padova models are considered the most complete models. Both models are based on clinical data sets instead of literature. The differences between the models are seen in the compartmental structures, the number of parameters, and the number of differential equations. Usually, more compartments make the model more complex. One common issue in all the models is that no model considers the effect of exercise on glucose dynamics explicitly. So these models cannot be applied to describe glucose dynamics in a free-living condition where blood glucose homeostasis is interrupted due to physical exercise. Besides this, simulated activity signals cannot be produced for building and testing ML-based models.

The development process of a physiological model considered in this thesis has been inspired by the physiological knowledge and architecture of the blood sugar regulation model by **R.O Foster** et.al. [18]. Though exercise dynamics has not been considered in this model, it has some distinguishing characteristics which are as below.

1. Five crucial plasma variables (Glucose, FFA, Lactate, Insulin, and Glucagon concentration) have been incorporated for modeling glucose homeostasis in respect of the normal human body.
2. The concentration of each plasma variable is the **aggregated result of rates** of some predefined release/uptake processes of a particular substrate. Each predefined release/uptake process has a basal rate in the basal condition.
3. The effect of concentration of a particular substrate (Glucose/FFA/Lactate/Insulin/ Glucagon) on the basal rate of any physiological process has been represented by a **static lookup table** of substrate concentration and its corresponding effect.
4. Time dimension has been considered as a collection of **discrete-time intervals** ($\Delta T = K - J$) to form the equations of the model. The equation of substrate (Glucose, FFA, Lactate, Insulin, and Glucagon) amount in a distribution volume at time point K is represented as.

$$A_K = A_J + \Delta T * (\sum R_{in} - \sum R_{out})_{JK};$$

$$C_K = A_K / (DistributionVolume);$$

Where,

A_K and A_J is the amount of substrate at time point K and J respectively;

C_K is the concentration of that substrate at time point K.

$\sum R_{in}$ means the sum of all inflow rates of a substrate.

$\sum R_{out}$ means the sum of all outflow rates of a substrate.

$(\sum R_{in} - \sum R_{out})_{JK}$ is the net change in rate (mg/min) of time interval JK

5. The rate of any metabolic process (uptake/release) is the multiplication of basal rate with effects of associated substrates retrieved from a static lookup table.

$$Rate = Basal_Rate * Effect(C_1) * Effect(C_2) * \dots * Effect(C_n);$$

$$Effect(C_n) = Static_Lookup_Table[Effect, Concentration];$$

Where, C_n is the concentration of n^{th} associated substrate.

6. For explaining the process of maintaining the energy requirement of the muscle and nervous system, in absence of or in low concentration of glucose, the role of Free Fatty Acid (FFA) production from the adipose tissue sector has been described quantitatively and explicitly.
7. A full model has been established based on human anatomy. Hence other sub-models such as simulating carbohydrate ingestion, injected insulin appearance in plasma and respiratory effect due to physical activity on glucose metabolism can be integrated as per physiology.

2.4 Mathematical models of glucose-exercise dynamics

Though the exercise dynamics has not been implemented in recognized glucose dynamics models/simulators till writing of this thesis there are some promising exercise dynamics models based on the minimal model of glucose dynamics available in the literature. A reliable physical activity model with glucose dynamics is essential especially for the model-based control design of Artificial Pancreas (AP) systems. Up to date, only a few physiological models are reported to describe how physical activity affects glucose dynamics.

Roy & Parker [13] proposed a minimal exercise model by extending Bergman's minimal model [25]. The model is prominent in the literature of exercise dynamics in blood glucose regulation and capable of describing the changes of basal insulin, glucose uptake, and endogenous glucose production during and after exercise. But, the long-lasting effect on insulin sensitivity is not considered in the model. Though the model can predict plasma glucose and insulin concentrations in response to mild-to-moderate exercise challenges, knowledge-driven lumped minimal models do not essentially resemble the glucoregulatory system at various levels of the body such as tissues and/or organs. To mimic the complex glucoregulatory system, comprehensive knowledge-driven models are required, which generally possess many parameters relating to the physiological process.

Breton [14] develops a mathematical model of glucose dynamics that can describe recognized changes in glucose regulation during physical exercise. Heart rate (beats per minute) is used to



Figure 2.8: Insulin Subsystem of S.M. Ewings model.



Figure 2.9: Insulin-dependent glucose subsystem of S.M. Ewings model.

detect and quantify the exercise intensity for the first time. The model can also describe both the short-term increase of insulin-independent glucose uptake and the prolonged insulin sensitivity change based on Bergman's minimal model [25]. The model was successfully fit to 21 type 1 diabetic subjects during a hyperinsulinemic clamp protocol and performance was compared with the standard minimal model of glucose kinetics that it was derived from. However, this model was designed following the parsimony principle and was kept intentionally small in terms of both states and parameters. Because of that, some important metabolic process such as the rise of Endogenous Glucose Production (EGP) caused by exercise is absent in the model.

Later on, **Dalla Man** et. al. [15] incorporated Breton's model into the comprehensive compartment model. Several modifications were made to reflect glucose dynamics with different exercise intensity and duration. Still, EGP increase is not modeled, and the model is validated with synthetic data of T1D patients [9]. The first attempt to describe the quantitative relationship between physical activity and change in insulin sensitivity has been taken in this model.

S.M. Ewings et. al. [16] presents a new statistical approach for analyzing the effects of everyday physical activity on blood glucose concentration in people with type 1 diabetes. This model shows that the exercise model based on minimal glucose dynamics cannot reflect the blood glucose concentration profile and returns non-viable results. The primary reason for poor model performance is the integration method of physical activity: a key assumption of the exercise model is that activity returns to basal level after exercise, but data from real dataset shows that physical activity does not return to basal level for extended periods during the day. This article also highlights important practical and theoretical issues not previously addressed in the quest for an artificial pancreas as a treatment for type-1 diabetes. The proposed method represents a new paradigm for the analysis of deterministic mathematical models of blood glucose concentration.

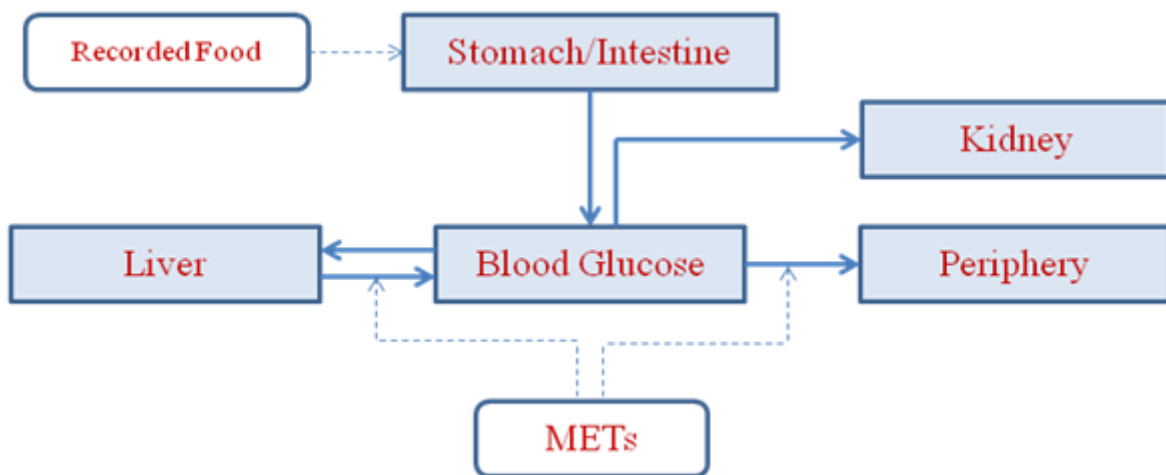


Figure 2.10: Insulin-independent glucose subsystem of S.M. Ewings model.

S.M.Ewings model consists of six (06) differential equations and eight (08) tunable parameters. Figure 2.8 – Figure 2.10 shows the different subsystems of the S.M. Ewings model. MET denotes metabolic equivalent of task. No endogenous production of insulin and glucagon has been assumed in the model. This means the model cannot be applicable for describing type-2 diabetes. The model is not capable of describing high-intensity exercise. Induction of insulin sensitivity due to physical exercise and its effect on the post-exercise period has been ignored. The model is not suitable for the long-term simulation of glycemic variation. S.M.Ewings model has been selected as a reference model for comparison with the proposed physiological model of this thesis.

M.C. Palumbo et. al. [26] formulated a novel computational system that is more detailed in describing both the physical exercise and the subjects' characteristics. This model has been built based on a previous model of whole-body metabolism by Kim et. al. [27]. Kim and colleagues proposed a whole-body, multi-scale computational model, incorporating cellular metabolism of different tissues/organs, to predict the responses of glucose, hormones (i.e., glucagon, insulin, epinephrine), and various substrates to moderate-intensity exercise. Exercise is described using the work rate (WR) expressed in Watt. A total of 136 differential equations consist of the multi-scale computational model. The modifications that have been made by M.C.Palumbo et. al. on the model by Kim are i) the use of the oxygen consumption in place of the WR (relative rather than absolute exercise intensity); ii) modeling how oxygen consumption relates to epinephrine; iii) explaining how the glucagon/insulin controller is modified by the new description of the exercise. To date, this model is the most comprehensive multi-scale exercise model available in the literature. But the model is computationally expensive for deployment in model predictive control in free-living conditions. This model is suitable for the development of diagnostic platforms for medical devices and patient-specific eHealth applications.

2.5 Research goal

Discussion and analysis on existing models presented in the previous two sections justify the research gap mentioned in section 1.3 and provide directions for designing integrated physiological model (IPM) of virtual diabetic patient. The main part of the IPM is the glucose dynamics in the blood circulation system. To implement the glucose dynamics of the IPM, glucose regulation model introduced by R.O.Foster [18] described at the end of section 2.3 can be adopted. The architecture of this model is suitable for introducing physiological constraints and exercise dynamics mathematically. For learning the physiology of exercise dynamics, existing models of section 2.4 are very important. The S.M.Ewings et.al [16] model can be treated as the state of the art and can be selected as a reference model for comparing the performance of the IPM. Once IPM implementation is complete, it will be usable as a standalone simulator in computer environment and applicable for OR on ML-based glucose forecasting. Hence, the implementation of the IPM as defined in section 1.4 is set as the research goal of this thesis.

Chapter 3

Methodology

According to the objectives of the thesis, it is required to build an ML model for producing forecasting of glucose concentration and a constraint-based glucose regulation model for OR on forecasting. Figure 1.5 showed a pictorial representation of OR on ML-based forecasting. In this chapter, adopted diabetic datasets and building strategy of the ML and OR models are described explicitly.

3.1 Datasets for modeling diabetes

Dataset is the fundamental component of any quantitative research. As it is already stated, blood glucose is primarily fluctuated due to ingestion of carbohydrate/glucose, external insulin injection, and physical activity in the daily life of a diabetic patient. Hence, a dataset consisting of those events and corresponding glycemic excursions is required to validate a physiological model. In the following sub-sections, four (04) datasets consisting of both clinical and free-living environments are described for getting illustration on the blood glucose fluctuation.

3.1.1 Oral glucose tolerance test (OGTT) data

A clinical dataset of OGTT performed by F.K. Knop et al. [28] was adopted. This dataset consisted of the mean values of plasma glucose and insulin concentration of all subjects at some specific time points of regular interval along the duration of test. In the test, ten (10) healthy subjects (eight men and two women) were selected, and a 50 g glucose tolerance test was performed. 17 blood samples were taken from the subjects during the test. Body weights of subjects were different from each other; hence all clinical data were scaled to a 70 kg body weight. The discrete values were interpolated using piecewise Hermite cubic interpolation in MATLAB to produce continuous series. Table B.1 in Appendix B contains the OGTT records.

3.1.2 Data on exercise in a clinical experiment

The clinical experiment conducted by G. Ahlborg and P. Felig [29] on physical activity was adopted in this research. In the experiment, 20 healthy non-obese adult male subjects were studied in the post-absorptive state after a 12 to 14-h over-night fast. For 3-4 days immediately before the exercise period, the subjects were told not to participate in any competitive athletics and to ingest meals consisting of 200-300 g of carbohydrate per day. The experiment was of 03-3.5 hours leg exercise (bicycle ergo-meter) with moderate intensity ($PVO_{2max} = 60\%$) and a 40-min post exercise recovery period. Data on oxygen consumption rate during exercise along with other substrates (Glucose, Insulin, Glucagon, FFA, and Lactate) concentration were recorded at regular interval for all subjects. From the mean value of recording of all subjects, piecewise Hermite cubic interpolation was used to produce a continuous signal. Table B.2 and Table B.3 in Appendix B contain the records of the exercise experiment.

3.1.3 CGM data of type-1 diabetic

To validate and compare the proposed model with a reference model, a single-day profile of CGM data of type-1 diabetes from the dataset shared by D.K. Rollins et al. [30] were used (Figure 3.1). The diet events and activity information of the selected profiles were transformed into the continuous signal by Elshoff et al. [31] and linear differential equation of oxygen consumption [13] respectively. The signal of insulin appearance into circulation was estimated from the exogenous insulin regimen using the model introduced by M. Berger, and D. Rodbard [32].

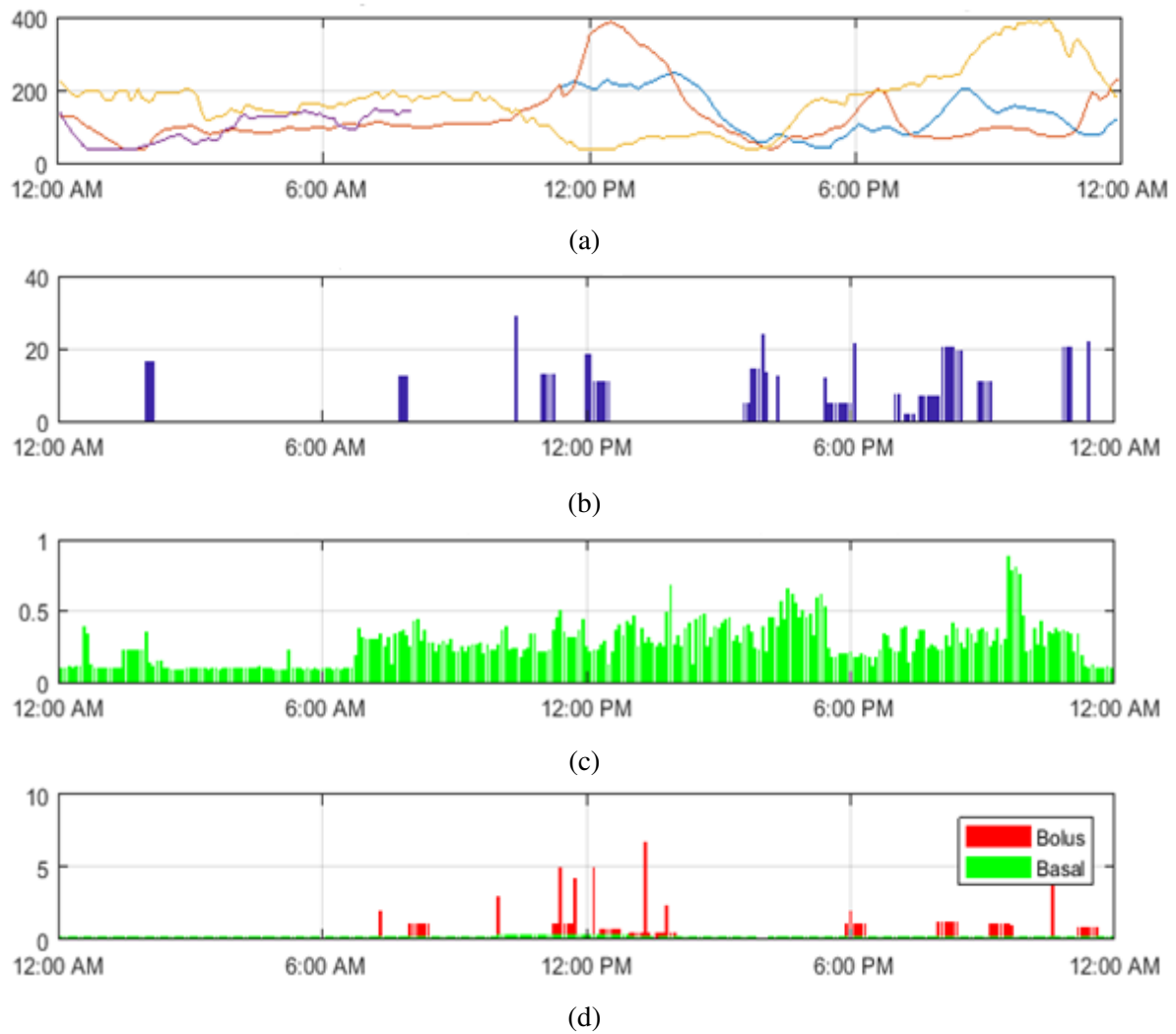


Figure 3.1: Dataset of a type-1 diabetic for 04 days: (a) CGM (b) Carbohydrate (c) Activity (d) Insulin.

3.1.4 CGM data of type-2 diabetic

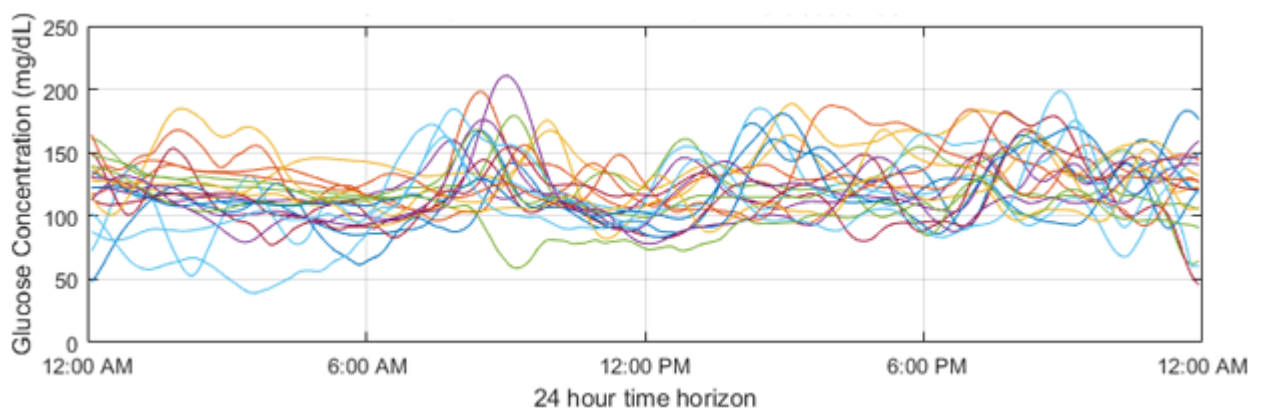


Figure 3.2: Daily pattern of glucose variation of a type-2 diabetic patient over CGM of 20 days.

A dataset of CGM, physical activity information, and diet events for 20 days of a type-2 diabetic patient was used in this research endeavor. Dataset was collected from a research publication by D.K. Rollins et. al. [30]. CGM data were smoothed by a MATLAB function and described over a 24 hour time horizon in Figure 3.2. From the time scale of the figure, it is observed that glucose fluctuation is steadier in the period of 12:00 am to 8:00 am of the day which is usually sleeping time. Outside of that time interval glucose concentration is more interrupted due to lifestyle events. Average of daily mean basal glucose is 114.24 mg/dl.

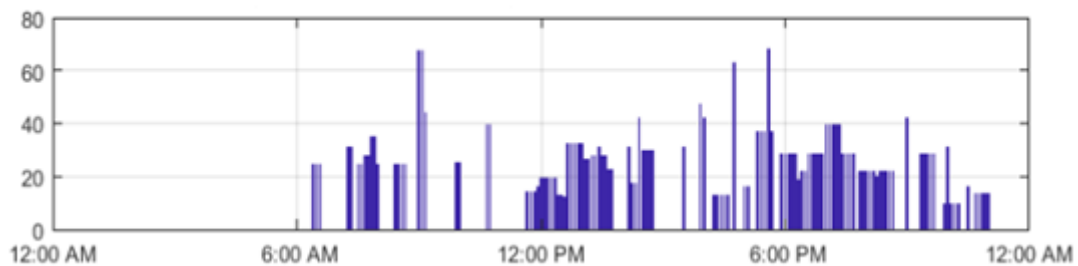


Figure 3.3: Daily pattern of carbohydrate(mg) taking in logged events over 20 days.

Carbohydrate (CHO) events shown in Figure 3.3 were logged as discrete events in the dataset. But the amount of CHO was transformed into a continuous physiological signal of glucose appearance (Figure 3.4) from the gut using a model by Elshoff et. al. [31].

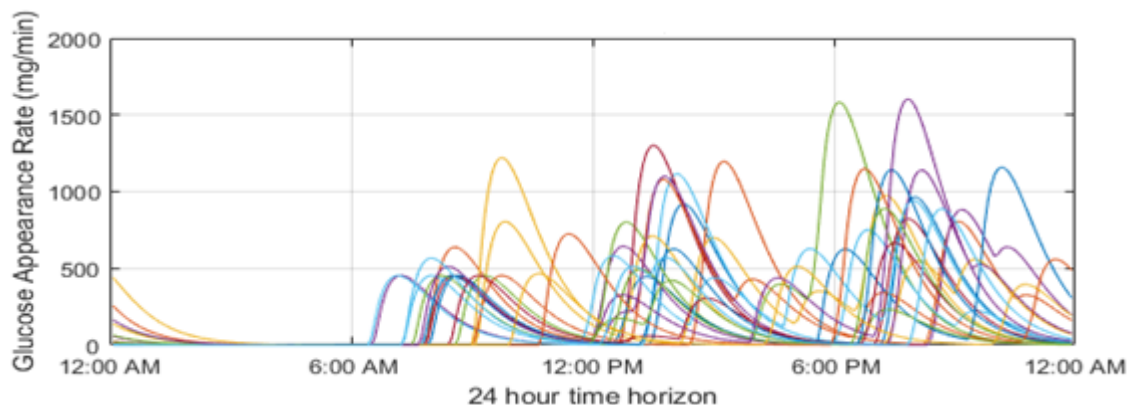


Figure 3.4: Continuous carbohydrate onboard (COB) profile of 20 days.

Figure 3.5 shows the correlation between glucose appearance rate from the gut and CGM signal. There is a strong positive correlation of about 45% among carbohydrate intake events and glucose variation.

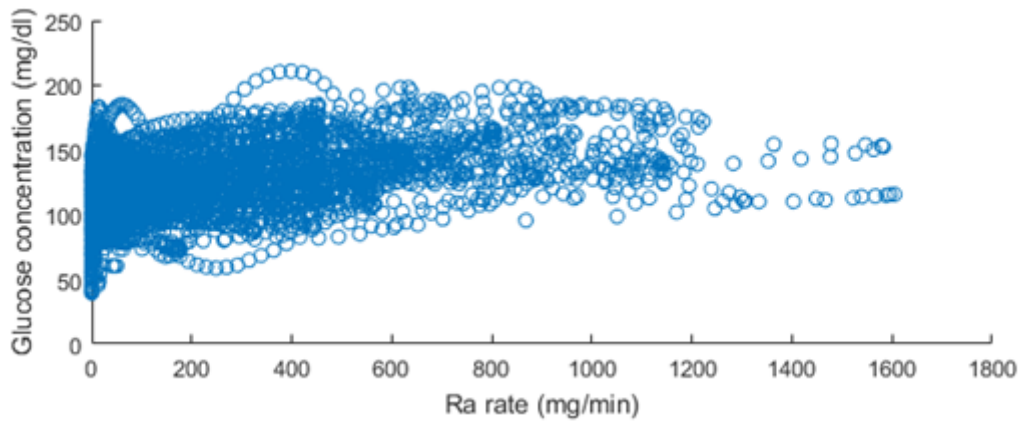


Figure 3.5: Correlation analysis of glucose appearance rate from the gut with glucose profile.

Figure 3.6 displays the continuous physical activity signal of a type-2 diabetic patient logged in free living environment. Physical activity signal maintains a negative correlation with CGM data and also has a very significant impact on blood glucose variation. But, unfortunately, Figure 3.7 shows a very poor correlation coefficient (-0.00896) of physical activity with the CGM data. The causes may be the incorporation of the noise factor introduced by the sensors.

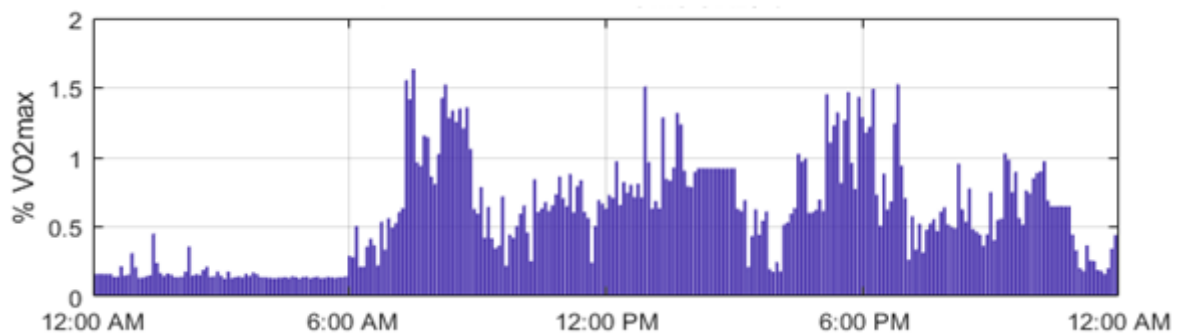


Figure 3.6: Physical activity profile of a type-2 diabetic patient over 20 days.

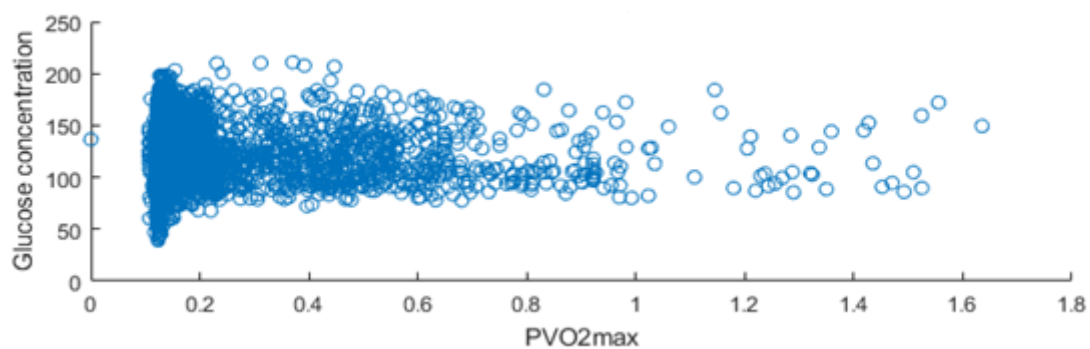


Figure 3.7: Correlation analysis of physical activity signal with glucose profile.

To observe the effect of correlation variation on the performance of the ML-based predictive model, a synthetic activity signal using a sigmoid function of CGM data has been generated as shown in Figure 3.8. The algorithm for generating synthetic activity signal is described in Appendix B. The synthetic activity has a good negative correlation of about 11.24% with the CGM data as displayed in the scatter plot in Figure 3.9. The utilization of synthetic activity signals doesn't influence the research result.

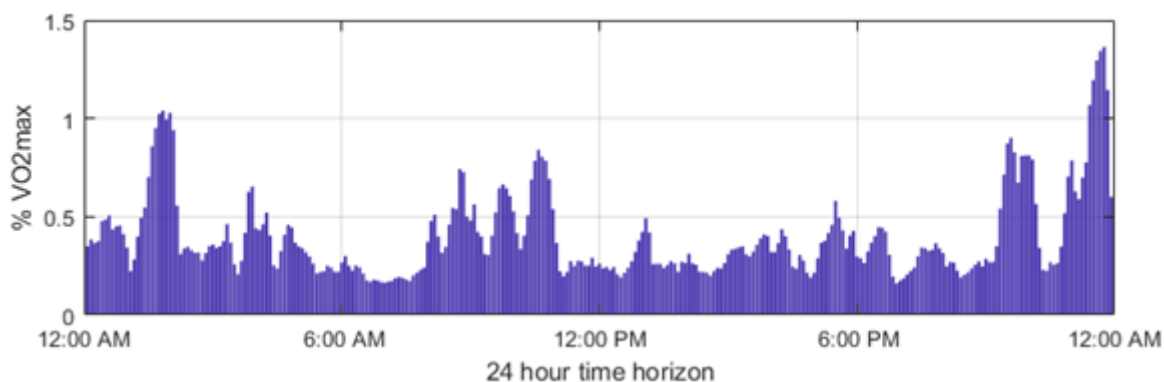


Figure 3.8: Synthetic activity profile based on CGM of a type-2 diabetic.

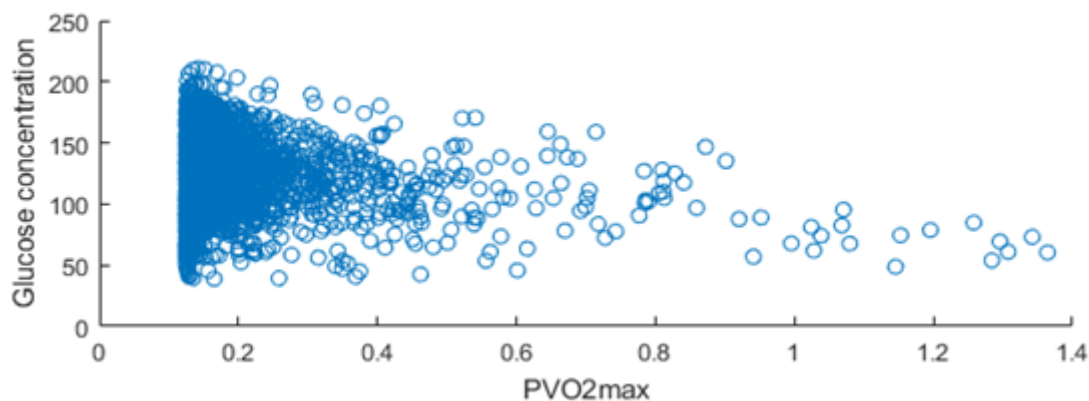


Figure 3.9: Correlation analysis of synthetic activity signal with glucose profile.

3.2 ML-based predictive model

According to the section 2.1.2, FFNN was mostly used for building predictive models for forecasting future glucose trajectories for a short period. Hence, in this experimentation of a new approach, FFNN was used to build an ML model using the CGM profile and logged information of a type-2 diabetic person described in section 3.1.4. Studying the structure of previous FFNN based models of the literature and analyzing correlation among the sequences of the diabetic dataset, the architecture of the ML model shown in Figure 3.10 is defined. The NN model

leveraged the correlation among glucose concentration data of time point $t+N$ with the glucose concentrations of time points t to $t-5$ and the accumulated amount of carbohydrate on board and activity intensity. Here N is the number of time points and it is determined by dividing the prediction horizon (PH) with sampling rate.

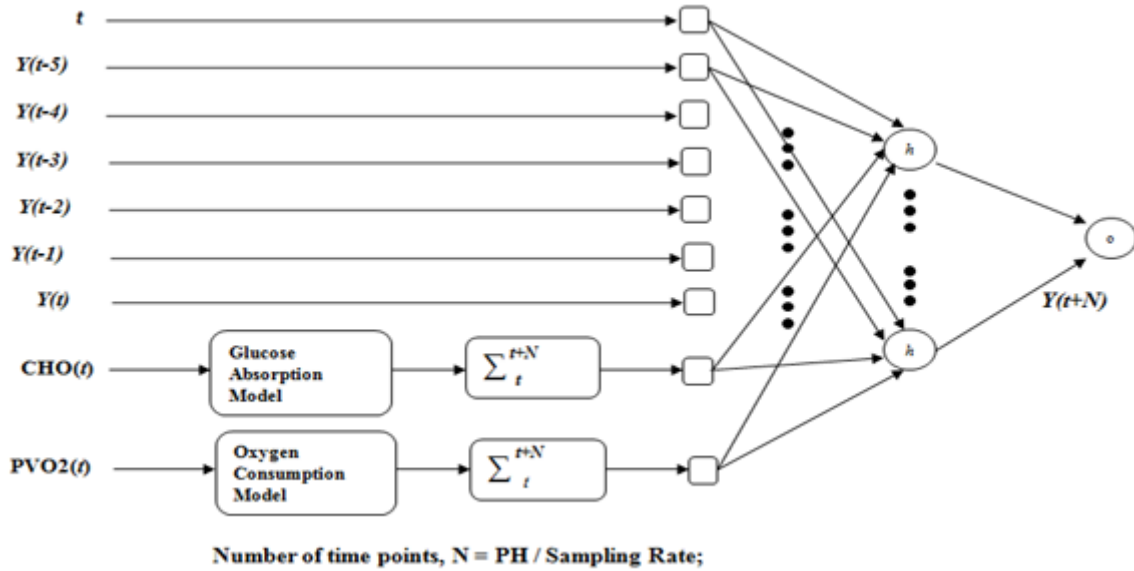


Figure 3.10: Architecture of FFNN based predictive model.

The model also took sample time points to consider the temporal variation of plasma glucose concentration over the daily life cycle. FFNN consisted of one hidden layer of ten (10) neurons that used hyperbolic tangent function as activation function and linear activation function in the output layer. The model was trained using a Bayesian regularization backpropagation algorithm with data of 15 days from the dataset and the rest of the records were used for model testing. Other parameters are the default setting of the *fitnet* function of the MATLAB Neural Network toolbox.

3.3 Physiology of blood glucose regulation

It is unavoidable to understand the pathophysiology of diabetes before diving into modeling mathematically. Existing research works [3]- [18] on mathematical modeling of blood glucose regulation are a good source of comprehensive knowledge on this topic. There are a couple of organs/tissues directly involved with the glucose metabolism in a human body listed in Figure 3.11. **Plasma** is the main carrier of various substrates (glucose, FFA and hormones) in the blood circulation system and only the accessible compartment in the human body. Since all organs/tissues are connected with the blood circulation system through vein and artery, substrates are released in or absorbed from plasma.

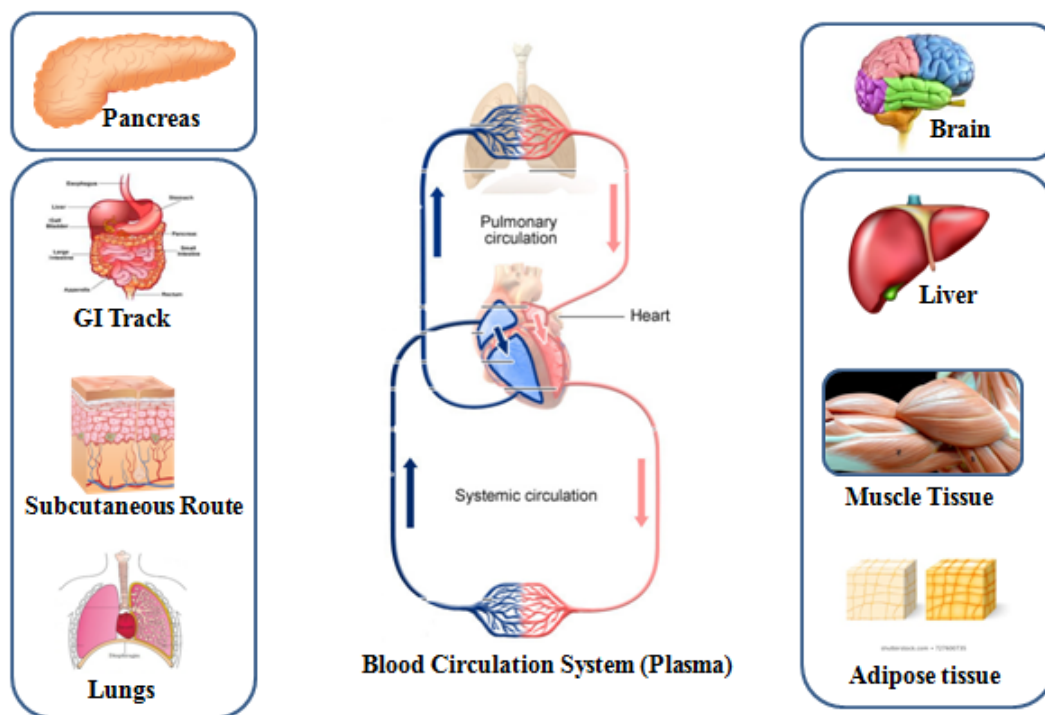


Figure 3.11: Organs/tissues involved in blood glucose regulation.

Glucose concentration, the integrated result of the net change of glucose in the plasma, is the main state variable that is needed to be maintained in the normal range in case of diabetes. In Figure 3.11, the **Pancreas** organ is responsible for secreting insulin and glucagon into plasma based on the concentration of glucose and other state variables. In diabetic conditions, the insulin secretion process is usually impaired or stopped permanently. The **brain** is the constant consumer of glucose and absorbs glucose without any dependency on insulin.

Liver, muscle, and adipose tissue are the three main crucial players in the blood glucose homeostasis condition of the body. The **liver** uptakes glucose for storing as glycogen and releases in plasma at a particular rate based on the concentration of plasma variables. Besides this, the liver converts free fatty acid and lactates into glycogen through gluconeogenesis and releases it into plasma for maintaining glucose homeostasis in fasting and intensive physical activity. **Muscle** tissues uptake glucose from plasma and burn it to produce energy for keeping the body active. It also stores glucose as glycogen for utilizing in demand and releases lactate into the circulation. **Adipose** tissues are responsible for releasing fatty acid in the circulation for maintaining basal secretion of insulin and the gluconeogenesis process of the liver. It also absorbs glucose from plasma and transforms it into fat for storing in the body.

Gastro-Intestinal (GI) Track and **Subcutaneous Route** are the paths via which oral ingestion of carbohydrate and injected insulin passes through respectively and flows into blood circulation. Orally ingested carbohydrate is degraded into glucose in the intestine and triggers the pancreas to secrete insulin in the circulation in advance. Insulin injection creates a flow of appearance in plasma based on type and dose of insulin. In the case of bolus insulin, the appearance rate is

high and within 15-30 minutes insulin dose is diffused completely in plasma. On the other hand, basal insulin takes a long time to diffuse into the blood from the injection site. **Lungs** consume oxygen from the air and transmit it into circulation to convey to the cells of the body. The rate of oxygen consumption is indirectly used for the measurement of activity intensity. The percentage of oxygen consumption usually rises to a higher rate during the onset of exercise and gradually comes down to a normal rate after the end of the exercise. Besides all of the organs/tissues mentioned above, the **kidney** also plays a role in blood glucose regulation. Glucose is excreted through urine if plasma concentration exceeds a threshold level.

In the following subsections, the behavior of various organs/tissues in terms of release/uptake of substrates against various events has been described by the dummy line graph over time dimension.

3.3.1 Fasting glucose homeostasis

During fasting, blood glucose concentration remains constant for a certain period. To maintain this constant situation hepatic release of glucose plays an important role in response to a higher secretion rate of glucagon from the **Pancreas**. When stored glycogen is reduced in a significant level, hepatic glucose release rate becomes low and plasma glucose concentration declines eventually as shown in Figure 3.12. To meet up the glucose requirement of **Brain** and **Muscle** tissue, the **Liver** increases the rate of the gluconeogenesis process to transform the fatty acid into glucose. But due to the slow rate of gluconeogenesis compare to glucose demand, blood glucose concentration falls ultimately. Simultaneously, the glucose uptake rate of **Muscle** and **Adipose** tissue is reduced.

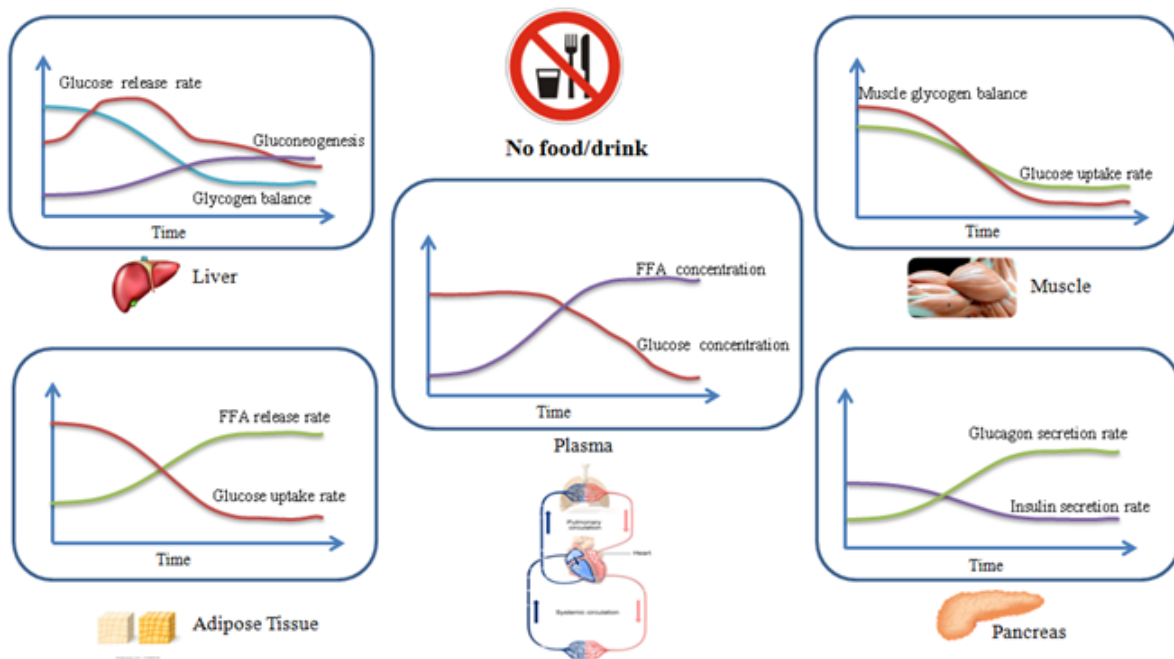


Figure 3.12: Body response during fasting.

3.3.2 Blood glucose regulation after carbohydrate ingestion

After consumption of food, glucose flux is produced by breaking down of carbohydrate at **Gut** and spreads out into circulation within 15-20 minutes of oral ingestion. Blood glucose concentration rises abruptly and **Pancreas** is stimulated to release insulin hormone in the circulation. Due to the higher concentration of insulin in plasma, glucose uptake rates in **Liver** and **Muscle** tissue are increased significantly to bring the glucose concentration to normal range. In the state of higher glucose concentration or hyperglycemia, hepatic glucose release rate is declined to the lower level and glycogen synthesis occurs. Figure 3.13 describes the response of the body after ingestion of carbohydrates.

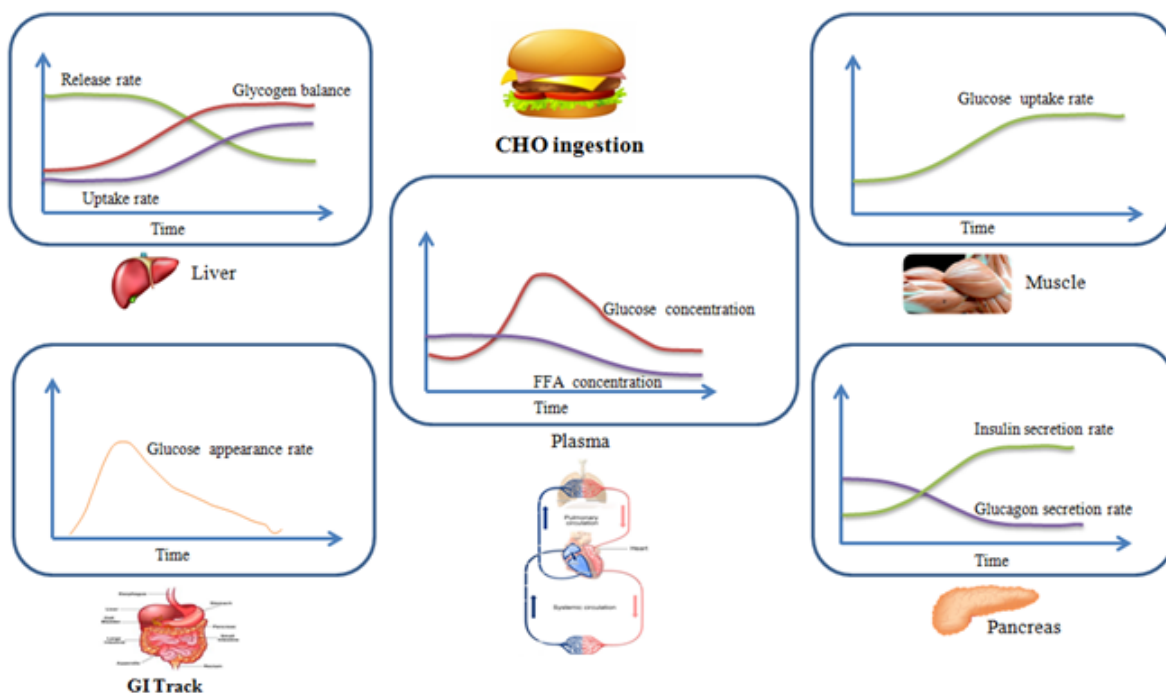


Figure 3.13: Body response after diet ingestion.

3.3.3 Blood glucose regulation after insulin injection

If insulin hormone is injected subcutaneously, it is transmitted into circulation at particular rate over time on the basis of insulin type and dose amount. Usually, insulin doses are of two types: bolus and basal. Bolus insulin, which has immediate action, is usually injected before a meal. Basal insulin, which is transmitted in blood circulation slowly, is injected for the long-term demand of insulin-dependent diabetic patients. A type-2 diabetic patient may have impaired insulin secretion from **Pancreas** or insulin resistance in **Peripheral** tissues and doesn't require external insulin in daily life. By performing regular physical exercise and a healthy diet routine glucose concentration can be maintained in the normal range. But for type-1 diabetic patient, who is unable to produce insulin completely, require essential external insulin injection at regular interval to keep blood glucose concentration in a safe range.

In a normal human body with the basal condition, blood glucose concentration falls immediately due to an increased rate of glucose uptake in **Peripheral** tissue after an injection of insulin dose causing hypoglycemia. To recover the normal condition, secretion of glucagon hormone from the **Pancreas** is increased to stimulate the **Liver** for releasing glucose in circulation by breaking down stored glycogen. Insulin secretion from **Pancreas** is suppressed completely in this condition. The description stated above has been illustrated in Figure 3.14.

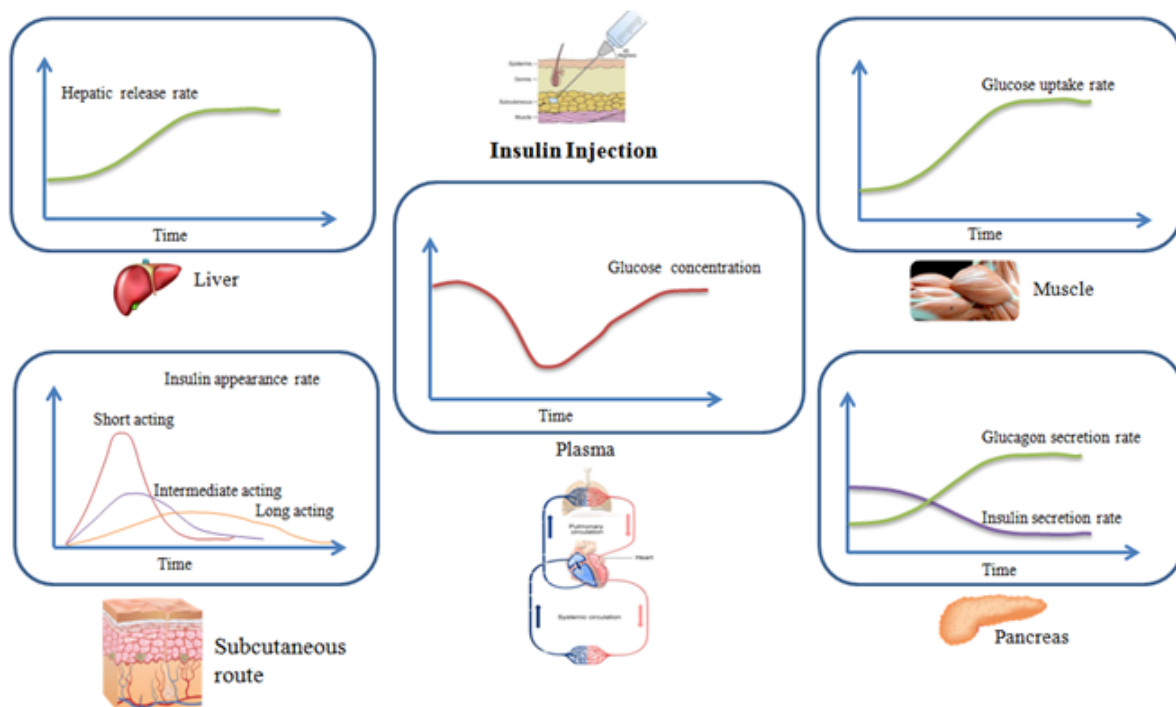


Figure 3.14: Body response after insulin injection.

3.3.4 Blood glucose regulation at physical activity

Physical exercise has severe short-term and long-term impacts on blood glucose regulation [33]. During physical activity, the rate of oxygen consumption increases from basal level to a constant extreme level within 4-5 minutes based on the intensity of the activity [13]. By integrating the changes in oxygen consumption rate it is possible to estimate the amount of effort which causes the disturbances in blood glucose concentration. Hence the rate of oxygen consumption is used as a metric to measure the intensity of physical activity. In some research work, the heart-bit rate is used to quantize the activity.

At the commencement of exercise with moderate intensity the **Skeletal Muscle** starts to increase the rate of uptake of glucose from the circulation. This increment does not produce any impact on the overall plasma glucose concentration. But as the duration of exercise prolongs, glucose concentration starts to fall gradually. The **Skeletal Muscle** tissue burns the glucose, taken from circulation, to meet the energy requirement of the body and releases lactate into the circulation. As per the metabolism of glucose homeostasis, glucagon hormone is released in the circulation for stimulating the **Liver** to release glucose to maintain the normoglycemia. But the amount of glycogen stored in the Liver isn't enough to compensate for the glucose demand for more than a few minutes. As the stored glycogen declined significantly Liver starts to transform non-glycogenic substrates such as fatty acid into glucose by gluconeogenesis process. But the rate of glucose production through gluconeogenesis is very slow in respect of the glucose requirement rate of the body during exercise. Eventually, the ultimate level of glucose concentration falls

and may cause hypoglycemia. In the case of exercise with high intensity, the sequence of events stated above occurs more rapidly. Type-1 or insulin-dependent diabetic patients can produce hypoglycemic conditions during exercise and may cause severe health conditions if a proper measure isn't taken on external insulin injection. Figure 3.15 describes the response of the body during the onset of exercise.

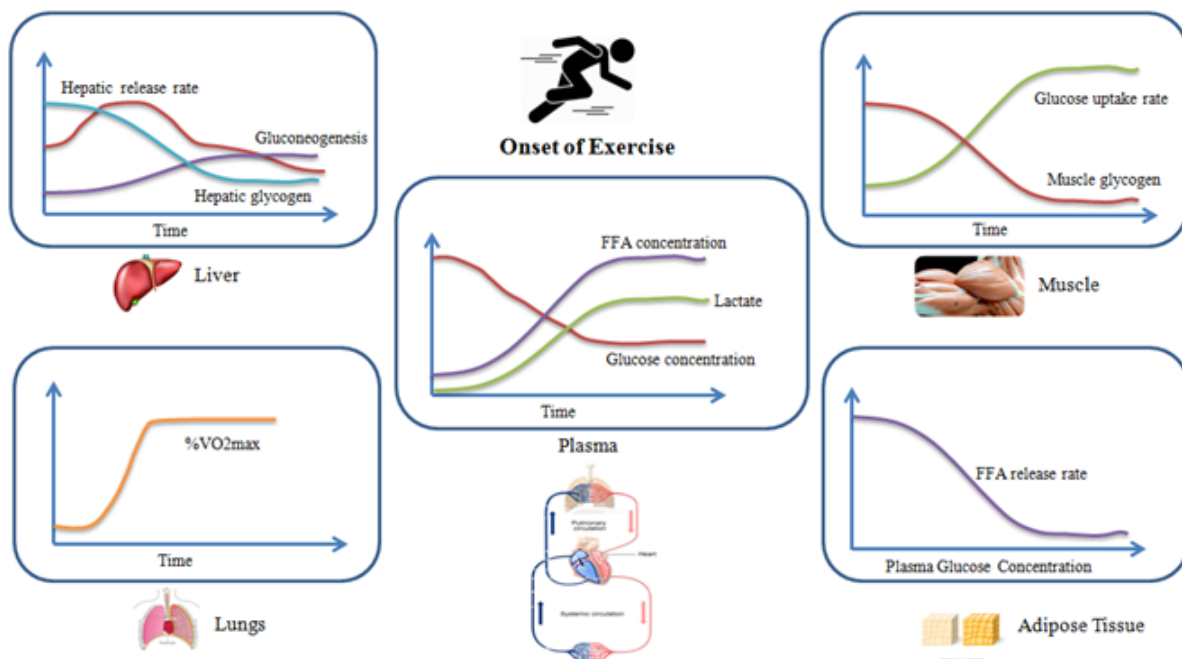


Figure 3.15: Body response during the onset of exercise.

At the end of the exercise, a separate metabolic process, called as recovery phase of the exercise, is started to bring the body into normal condition. In the case of exercise of low intensity, the recovery phase doesn't produce any significant changes in the circulation. But at the end of prolonged exercise, the glucose uptake rate of the **Skeletal Muscle** is declined abruptly and at the same time, the gluconeogenesis rate is increased rapidly in presence of a higher concentration of lactate. This lactate is released from the Skeletal Muscle during exercise. Since glucose reproduction is greater than the uptake rate, plasma glucose concentration rises gradually and returns to the normal range within a certain period. In the case of impaired liver function, glucose homeostasis conditions may not be achieved. Hence precaution should be taken before attempting high-intensity exercise for diabetic patients. Recovery phenomena have been described in Figure 3.16.

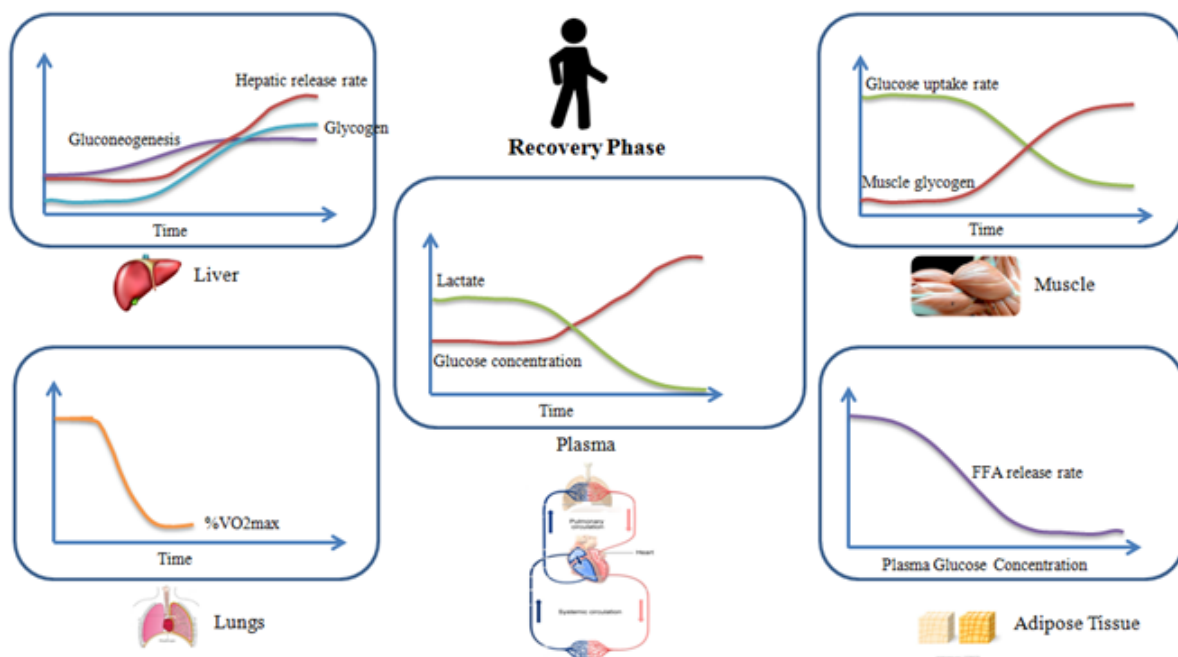


Figure 3.16: Body response just after exercise.

Besides the short-term effect of physical exercise, there is also a long-term impact on insulin sensitivity in peripheral tissue of the body. Usually, after prolonged exercise, the human body gains an extra capability on the glucose uptake process in **Muscle** tissue. In this case, the metabolic process of glucose uptake without the influence of insulin gets an accelerated rate which ultimately helps to prevent the hyperglycemic condition in the body. This accelerated rate may persist for 10-12 hours after exercise and may prevalent on basis of the regularity of the exercise schedule [33].

3.3.5 Rate-balance-concentration paradigm

By observing the metabolic behavior of organs/tissues discussed so far (section 3.3.1-3.3.4), a relation among the rate of metabolic processes (release/uptake), balance of plasma substrate, and concentration has become obvious to describe the glucose dynamics. The **rate** of metabolic processes (release/uptake) is the product of their basal rate and effects of single or multiple state variables. The **balance** of a specific substrate in the whole body is the result of the integration of net change in rates of corresponding metabolic processes. The **concentration** of a specific substrate is the amount of substrate balance per unit distribution volume of the body. Due to external/internal disturbances, any change in concentration eventually makes an effect on the rate of associated metabolic processes. So, the dynamics of the blood glucose regulation makes a **rate-balance-concentration** paradigm (Figure 3.17).

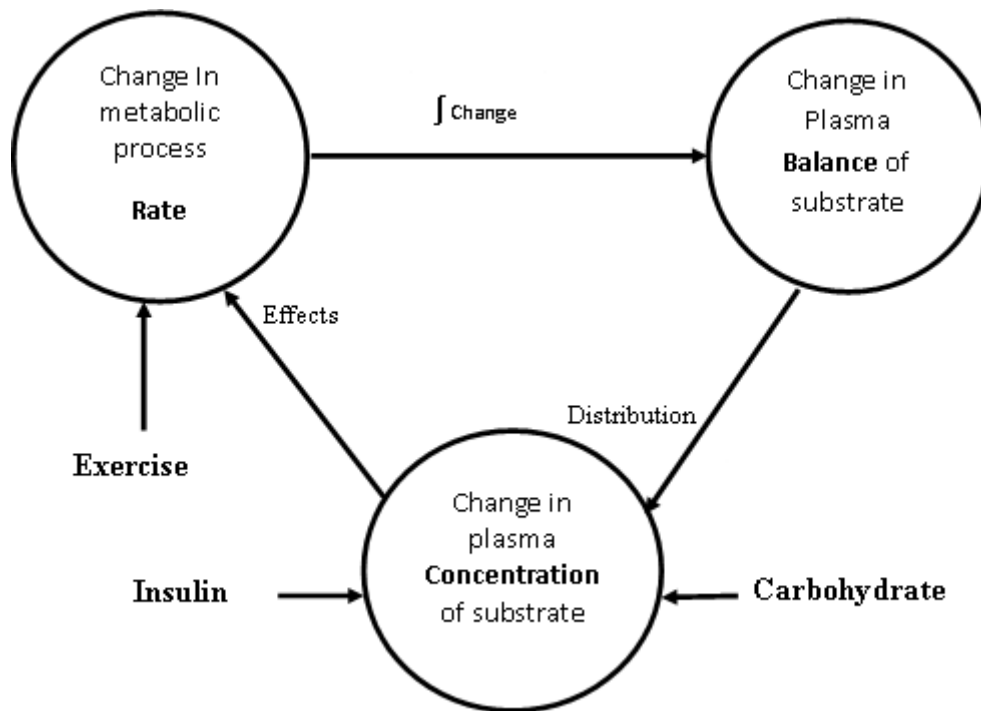


Figure 3.17: Rate-balance-concentration paradigm in blood glucose regulation.

In response to external stimuli such as oral ingestion of meals or injection of insulin or physical exercise, the **rate-balance-concentration** paradigm negotiates the disturbances by altering the metabolic process rate to bring the stable condition in the body. Both oral ingestion of carbohydrate and subcutaneous injection of insulin directly interrupt plasma concentration. On the other hand, exercise intensity has an immediate impact on the metabolic process rate of glucose uptake and lactate release by the skeletal muscle tissue.

3.3.6 Defining dynamics mathematically

As described by Charette et. al. [17], the effect of any substrate concentration/external stimulus on the basal rate of a metabolic unit process can be modeled using the kinetic theory of enzyme-catalyzed chemical reactions. According to Charette et. al., a plot of the rate of product formation as a function of substrate concentration with enzyme activity as a parameter is “sigmoid” in shape, i.e. the process is rate-limited at some saturation value (Figure 3.18).

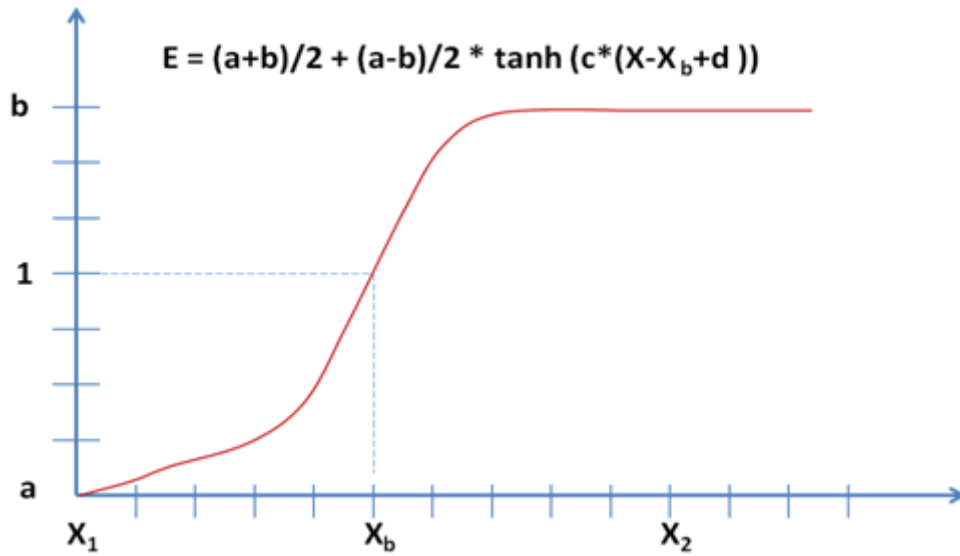


Figure 3.18: Hyperbolic tangent function for effect representation.

For this reason, to represent the effect of concentration of any substrate/external-stimulus/event on the basal rate of any metabolic unit process involved in blood glucose regulation in the human body, the following functional form can be applied.

$$E_x = E_1 + E_2 * \tanh(c * (X - X_b + d))$$

where,

$$E_1 = (a + b)/2$$

$$E_2 = (a - b)/2$$

X = Substrates concentration/Stimulus magnitude (input);

X_b = Basal concentration/magnitude of any substrate/stimulus;

E_x = Effect of fluctuation of Substrates/Stimulus (output);

a = Effect at lowest point of X ;

b = Effect at highest point of X ;

c = Slope of effect trajectory;

d = Adjustment offset;

In the proposed physiological model of this thesis, the above form of hyperbolic tangent function has been applied to represent the effect of change of state variables on various metabolic processes of glucose dynamics. It has been considered that any metabolic process rate is

increased monotonically from some nominal value until saturation occurs under the appropriate stimulus. If any metabolic process is stimulated by multiple substrates concentration then the basal rate of that process shall be multiplied by the effect of each substrate to obtain the ultimate metabolic rate as below.

$$\text{MetabolicRate} = \text{BasalRate} * E_1 * E_2 * \dots * E_n;$$

Where, n is the number of substrates.

There were two fundamental objectives for adopting the form of hyperbolic tangent function in the modeling rate-balance-concentration paradigm. The first one was the requirement of the physiological model for saturation constraints. The next one was to get a metabolic spectrum of effects of plasma variables on metabolic processes for physiological interpretation.

Chapter 4

Integrated Physiological Model

As mentioned in section 1.4 and illustrated in the methodology, equations of the IPM of virtual diabetic patient are described explicitly in this chapter. Models are implemented in Matlab/Simulink environment (Appendix C) and results of various fitting experiments are mentioned and discussed in the following chapters.

4.1 Constraint-based glucose regulation

Considering five (05) plasma variables, sixteen (16) rates of metabolic processes in different organs/tissues, and hyperbolic tangent function of saturation relation of biochemical reactions, the constraint-based physiological model is formed according to the physiological architecture presented in [18] and discussed in section 2.3. There are three secondary state variables in the model: storage of glycogen in the liver and peripheral tissues, and insulin sensitivity induced due to physical activity.

4.1.1 Glucose dynamics

Glucose release from liver: $HPT_Release_Rate = Basal_Rate \times E_{glucose} \times E_{insulin} \times E_{glucagon} \times E_{glycogen}$;

Glucose uptake by liver: $HPT_Uptake_Rate = Basal_Rate \times E_{glucose} \times E_{insulin} \times E_{glycogen}$;

Glucose uptake by muscle tissue: $MT_Uptake_Rate = Basal_Rate \times E_{glucose} \times E_{insulin} \times E_{peri_store} \times E_{exercise}$;

Glucose uptake by adipose tissue: $APT_Uptake_Rate = Basal_Rate \times E_{glucose} \times E_{insulin}$;

Since effect of insulin concentration on glucose uptake process is inversely influenced by concentration of FFA, $E_{insulin} = (a + b)/2 + (a - b)/2 * \tanh(c * ((Insulin * E_{FFA}) - Basal + d))$;

Glucose uptake by nervous system: $NS_Uptake_Rate = Basal_Rate \times E_{glucose}$;

Glucose uptake by red blood cell: $RBC_Uptake_Rate = Basal_Rate \times E_{glucose}$;

Glucose extraction through urination: $Urine_Spillage_Rate = 0.5 \times (glucose - threshold)$;

$NetChange = GUT_Appearance_Rate + HPT_Release_Rate - HPT_Uptake_Rate$

$$-MT_Uptake_Rate - APT_Uptake_Rate - NS_Uptake_Rate \\ -Urine_Spillage_Rate - RBC_Uptake_Rate;$$

GUT_Appearance_Rate denotes the rate of glucose appearance in circulation from the gut after oral ingestion. $PlasmaGlucoseStorage = InitialPlasmaGlucoseLevel + \int NetChange;$

$$PlasmaGlucoseConcentration = PlasmaGlucoseStorage / (VolumeofDistribution * 10);$$

$$VolumeofDistribution = 20\% * (BodyWeight);$$

Lactate uptake by liver: $HPT_Lactate_Uptake_Rate = Basal_Rate \times E_{Lactate} \times E_{glycogen} \times E_{exercise};$

$$NetChange = HPT_Uptake_Rate - HPT_Release_Rate + HPT_Gluconeogenesis_Rate + \\ HPT_Lactate_uptake_rate;$$

$$HepaticStorage = InitialHepaticGlucoseLevel + \int NetChange;$$

$$HepaticGlycogen(\%) = HepaticStorage / (HepaticStorage + LiverWeight);$$

4.1.2 Exercise dynamics

Estimation of effort during exercise: $Effort = \int_{StartTime}^{EndTime} Oxygen_Consumption_Rate$

Induction of insulin sensitivity (IS) for given effort:

$$IS_Induction_Factor = (a + b)/2 + (a - b)/2 * \tanh(c * (Effort - Basal + d));$$

The net result of insulin sensitivity: $Net_IS_Factor = \int (IS_Induction_Factor - IS_Degradation_factor);$

Glucose burning in muscle tissue: $MT_Utilization_Rate = Basal_Rate \times E_{peri_store} \times (E_{exercise} + Net_IS_Factor);$

Lactate releasing from muscle tissue: $MT_Lactate_Release_Rate = Basal_Rate \times E_{peri_store} \times E_{exercise};$

$$NetChange = MT_Uptake_Rate - MT_Utilization_Rate - MT_Lactate_Release_Rate;$$

$$PeripheralGlucoseStorage = InitialPeripheralGlucoseStorage + \int NetChange;$$

4.1.3 Free Fatty Acid (FFA) dynamics

FFA release from adipose/fat tissue: $FFA_Production_Rate = Basal_Rate \times E_{glucose} \times E_{insulin};$

FFA transformation in liver: $HPT_Gluconeogenesis_Rate = Basal_Rate \times E_{FFA} \times E_{glycogen};$

FFA utilization in plasma: $FFA_Utilization_Rate = ((TotalPlasmaFFAAmount * E_{FFA}) / TCFAU);$

Where, TCFAU = Time Constant for FFA Utilization in fasting = $1.443 * half_life = 6;$

$$NetChange = FFA_Production_Rate - HPT_Gluconeogenesis_Rate - FFA_Utilization_Rate;$$

$$TotalPlasmaFFAAmount = InitialPlasmaFFAlevel + \int NetChange$$

$$FFAConcentration = PlasmaFFAStorage / (VolumeofDistribution * 150);$$

$$VolumeofDistribution = 10\% * (BodyWeight);$$

4.1.4 Lactate dynamics

$$NetChange = MT_Lactate_Release_Rate - HPT_Lactate_Uptake_Rate;$$

$$TotalPlasmaLactateAmount = InitialPlasmaLactateLevel + \int NetChange$$

$$PlasmaLactateConcentration = TotalLactate / VolumeofDistribution$$

$$VolumeofDistribution = 20\% * (BodyWeight);$$

4.1.5 Insulin dynamics

$$\text{Insulin secretion from pancreas: } I_Secretion_Rate = Basal_Rate \times E_{change} \times E_{glucose} \times E_{FFA};$$

$$E_{change} = (a + b)/2 + (a - b)/2 * \tanh(c * (perceived_glucose_change - Basal + d));$$

$$\text{Insulin degradation in plasma: } I_Degradation_Rate = 0.075 \times (NetInsulinAmount) \times E_{exercise};$$

$I_Appearance_Rate$ denotes the rate of insulin appearance from the site of subcutaneous injection.

$$NetChange = I_Secretion_Rate - I_Degradation_Rate + I_Appearance_Rate;$$

$$TotalInsulin = InitialPlasmaInsulinLevel + \int NetChange;$$

$$PlasmaInsulinConcentration = TotalInsulin / VolumeofDistribution$$

$$VolumeofDistribution = 20\% * (BodyWeight);$$

Advance sensing of plasma glucose concentration by pancreas is represented as below.

$$perceived_glucose_concentration = PlasmaGlucoseLevel + GUT_Appearance_Rate * 0.20;$$

$$Perceivedglucosechange = (perceived_glucose_concentration(t) - perceived_glucose_concentration(t-1)) / 0.1;$$

Where t is the timestamp.

4.1.6 Glucagon dynamics

$$\text{Glucagon secretion from pancreas: } G_Secretion_Rate = Basal_Rate \times E_{glucose};$$

$$\text{Glucagon degradation in plasma: } G_Degradation_Rate = 0.20 \times (NetPlasmaGlucagonAmount);$$

$$NetChange = G_Secretion_Rate - G_Degradation_Rate;$$

$$TotalGlucagon = InitialPlasmaGlucagonLevel + \int NetChange;$$

$$PlasmaGlucagonConcentration = TotalGlucagon / VolumeofDistribution.$$

$$VolumeofDistribution = 20\% * (BodyWeight);$$

4.1.7 Modeling external stimulus

Guts glucose appearance: After having diet glucose is appeared in blood circulation from the guts breaking down the carbohydrate part of the diet at a consistent rate based on the amount of ingested carbohydrate, type of carbohydrate, and elapsed time. The function of guts glucose appearance is modeled in several studies. In this research, the model proposed by Elashoff et al. [31] is used to produce a glucose appearance signal. ‘GUT_Appearance_Rate’ variable in the model represents the rate of glucose appearance from the gut.

Subcutaneous insulin appearance: Similarly, insulin appearance in circulation from the site of subcutaneous injection is dependent on the amount of dose, insulin type, and time. The model of subcutaneous insulin diffusion proposed by M. Berger, and D. Rodbard [32] is used to produce plasma insulin appearance signals from injected dose in this thesis. The ‘I_Appearance_Rate’ variable in the proposed physiological model represents the rate of insulin appearance from the exogenous insulin injection.

Oxygen consumption rate: The intensity of exercise is measured with the oxygen consumption rate of the body. There is a delay between the onset of exercise and the physiological response of the body. That delay is needed to be modeled for incorporating exercise dynamics with the glucose regulation. The equations introduced in A. Roy and R.S. Parker [13] for modeling physiological delay are adopted in this thesis. The ‘Oxygen_Consumption_Rate’ variable in the proposed physiological model represents the continuous signal of oxygen consumption during the onset of exercise.

All three models: [31], [32], and [13] are described in Appendix A. ‘Basal_Rate’ in all equations represented the basal rate of the corresponding metabolic process. Values of basal rates of various metabolic processes and constants were taken from the literature [18]. Because of having a secondary role of FFA and lactate in glucose regulation, glucose dynamics can be modeled considering the concentration of glucose, insulin, and glucagon only.

4.2 Parameter estimation by model optimization

Parameter constraints: Each effect equation has four (04) parameters to represent the saturation phenomenon. So, the number of total tunable parameters depends on the number of total effect equations defined in the model. But using the physiological knowledge and experiences of several trials on optimization it appeared that several parameters could be easily assumed without going into the optimization process. One of those parameters was the lowest magnitude of effect for the concentration of a particular plasma variable on a particular process rate. It may be the parameter ‘a’ or ‘b’ in the hyperbolic tangent effect equation (Figure 3.18). Another parameter is ‘d’ which was used to set the midpoint of the operating range of concentration of plasma variables. Hence keeping those assumable parameters aside only half of the total parameters were needed to tune by optimization over the selected experimental dataset. The selected parameters were the highest magnitude of the effect equations (‘a’ or ‘b’) and the slope of the effect trajectories (‘c’ parameter). In the parameter estimation process, an iterative procedure was used, which used a sequential quadratic programming (SQP) method for solving the nonlinearly constrained optimization problem. Basal and initial states of plasma variables were set before starting the optimization process based on the experimental dataset and assumption. The constraints imposed on the model parameters during the optimization were following.

1. Each effect equation was equal to unity in the case of basal concentration of the corresponding substrate. i.e. $(a + b)/2 + ((a-b)/2) * \tanh(c * d) = 1$
2. Parameters 'a' and 'b' were positive in each case. The range of parameters of higher magnitude ('a' or 'b') was set to $1 \rightarrow Inf$ and parameter of lower magnitude was assumed with physiological knowledge.
3. Parameter 'c' was always negative in each effect equation and set to a range achieved by a lot of trial and error.
4. $X_m = [X_b - (\pm d)]$ indicated the midpoint of the operating range of the substrate concentration where X_b was the basal concentration. Hence the parameter 'd' was such that X_m didn't exceed the interval $[X1, X2]$.

The initial value of the parameters plays a very important role in the optimization and these values are obtained from the experience of a huge number of trial optimization.

Objective function: The objective function of the optimization problem is to minimize the sum of deviation of model results from the clinical data. The best model parameters result in the closest model results to the clinical data. In the case of the OGTT experiment with dataset described in 3.1.1, the objective function of the constrained nonlinear optimization problem is the following.

$$\min_{\Theta} \sum_{i=1}^{i=n} (|G_m^i - G_c^i| + |I_m^i - I_c^i|)$$

Where G_m^i and I_m^i are plasma glucose and insulin concentrations at the time i obtained from the model, respectively; G_c^i and I_c^i are corresponding clinical measurements. n is the size of the clinical OGTT dataset; Θ is the vector of selected model parameters. In the case of the exercise experiment with dataset described in 3.1.2, the objective function of the constrained nonlinear optimization problem is the following.

$$\min_{\Theta} \sum_{i=1}^{i=n} (|G_m^i - G_c^i| + |F_m^i - F_c^i| + |L_m^i - L_c^i| + |I_m^i - I_c^i| + |\Gamma_m^i - \Gamma_c^i|)$$

Where $G_m^i, F_m^i, L_m^i, I_m^i$ and Γ_m^i are plasma glucose, free fatty acid, lactate, insulin, and glucagon concentrations at the time i obtained from the model, respectively; $G_c^i, F_c^i, L_c^i, I_c^i$ and Γ_c^i are corresponding clinical measurements. n is the size of the clinical exercise dataset; Θ is the vector of selected model parameters. In the case of the CGM experiment with dataset described in 3.1.3 and 3.1.4, the objective function of the constrained nonlinear optimization problem is the following.

$$\min_{\Theta} \sum_{i=1}^{i=n} (|G_m^i - G_{cgm}^i|)$$

Where G_m^i is plasma glucose concentrations at the time i obtained from the model; G_{cgm}^i is corresponding clinical measurements. n is the size of the CGM dataset; Θ is the vector of selected model parameters.

4.3 Reshaping physiological model for OR

CGM profile of a type-2 diabetic along with diet and exercise log is selected for OR on forecasting obtained from FFNN. Hence physiological model presented in section 4.2 is needed to reshape with less number of state variables. The reshaped model can also be used for fitting other experimental datasets which consist of a limited number of plasma variables (section 5.1, 5.3, 5.4, and 5.7). Table 4.1 shows the tabular representation of the reshaped physiological architecture of the adopted model.

Table 4.1: Tabular representation of metabolic relation among plasma variables and metabolic processes.

Plasma Variables	Liver		Muscle	Pancrease		Plasma		
	Glucose Release	Glucose Uptake	Glucose Uptake	Insulin Secretion	Glucagon Secretion	Insulin Degradation	Nerve Uptake	RBC Uptake
Glucose								
Insulin								
Glucagon								
Exercise								

* dashed line indicates the neutral effect of corresponding plasma variables on metabolic processes.

The mathematical architecture of the reshaped model according to Table 4.1 is described below.

4.3.1 Glucose dynamics

Glucose release from liver: $HPT_Release_Rate = Basal_Rate \times E_{glucose} \times E_{insulin} \times E_{glucagon}$;

Glucose uptake by liver: $HPT_Uptake_Rate = Basal_Rate \times E_{glucose} \times E_{insulin}$;

Glucose uptake by muscle tissue: $MT_Uptake_Rate = Basal_Rate \times E_{glucose} \times E_{insulin} \times E_{exercise}$;

Glucose uptake by nervous system: $NS_Uptake_Rate = Basal_Rate \times E_{glucose}$;

Glucose uptake by red blood cell: $RBC_Uptake_Rate = Basal_Rate \times E_{glucose}$;

Glucose extraction through urination: $Urine_Spillage_Rate = 0.5 \times (glucose - threshold)$;

$$\begin{aligned} NetChange &= GUT_Appearance_Rate + HPT_Release_Rate - HPT_Uptake_Rate \\ &\quad - MT_Uptake_Rate - NS_Uptake_Rate - Urine_Spillage_Rate - RBC_Uptake_Rate; \end{aligned}$$

'GUT_Appearance_Rate' denoted the rate of glucose appearance in circulation from the gut after oral ingestion.

$$\begin{aligned} PlasmaGlucoseStorage &= InitialPlasmaGlucoseLevel + \int NetChange \\ PlasmaGlucoseConcentration &= PlasmaGlucoseStorage / (VolumeofDistribution * 10); \\ VolumeofDistribution &= 20\% * (BodyWeight); \end{aligned}$$

4.3.2 Insulin dynamics

Insulin secretion from pancreas: $I_Secretion_Rate = Basal_Rate \times E_{change} \times E_{glucose}$;

$$E_{change} = (a + b)/2 + (a - b)/2 * \tanh(c * (perceived_glucose_change - Basal + d));$$

Insulin degradation in plasma: $I_Degradation_Rate = 0.075 \times (NetInsulinAmount) \times E_{exercise}$;

'I_Appearance_Rate' denotes the rate of insulin appearance from the site of subcutaneous injection.

$$\begin{aligned} NetChange &= I_Secretion_Rate - I_Degradation_Rate + I_Appearance_Rate; \\ TotalInsulin &= InitialPlasmaInsulinLevel + \int NetChange; \\ PlasmaInsulinConcentration &= TotalInsulin / VolumeofDistribution. \\ VolumeofDistribution &= 20\% * (BodyWeight); \end{aligned}$$

Advance sensing of plasma glucose concentration by the pancreas is represented as below.

$$\begin{aligned} perceived_glucose_concentration &= PlasmaGlucoseLevel + GUT_Appearance_Rate * 0.20; \\ Perceivedglucosechange &= (perceived_glucose_concentration(t) - perceived_glucose_concentration(t-1)) / 0.1; \end{aligned}$$

Where, t is the timestamp.

4.3.3 Glucagon dynamics

Glucagon secretion from pancreas: $G_Secretion_Rate = Basal_Rate \times E_{glucose}$;

Glucagon degradation in plasma: $G_Degradation_Rate = 0.20 \times (NetPlasmaGlucagonAmount)$;

$$\begin{aligned} NetChange &= G_Secretion_Rate - G_Degradation_Rate; \\ TotalGlucagon &= InitialPlasmaGlucagonLevel + \int NetChange; \\ PlasmaGlucagonConcentration &= TotalGlucagon / VolumeofDistribution. \\ VolumeofDistribution &= 20\% * (BodyWeight); \end{aligned}$$

4.3.4 Estimating parameters and constraints range

After reshaping the proposed glucose regulation model for OR, the dynamics were represented by fourteen (14) effect equations. Hence there were fifty-six (56) tunable parameters present in

the model and only twenty-four (24) parameters were selected for tuning by optimization over the dataset of a type-2 diabetic according to the physiological representation of Table 4.1. There were two use cases of the constrained nonlinear optimization process according to the illustration of Figure 1.5. In the **first** case, optimization was conducted on original retrospective data to generate the highest magnitude of each effect equation ('a' or 'b') over the diurnal interval using parameter constraints described in section 4.2. By making an average of all magnitudes for a particular equation a mean magnitude was determined for the parameter of highest magnitude ('a' or 'b') of all effect equations. This set of magnitude was treated as personalized constraints. In the **second** case, optimization was conducted on the forecasted glucose profile of the ML model using similar parameter constraints of section 4.2 except the highest magnitude of all effect equations. The highest magnitude of all effect equations, which is **Inf** by default, is replaced with the personalized constraint estimated in the first case before optimization. After completion of each optimization, plotting of each effect equations gave the metabolic spectrum for the time interval of the day taken under consideration for optimization.

Chapter 5

Experimental Results and Discussion

For validating the proposed constraint-based physiological model described in the previous chapter, it is required to test the model upon various real world experiments. This chapter presents the experimental results conducted with dataset described in Chapter 3 and discusses the characteristics of the model responses.

5.1 Fitting OGTT experiment

After oral ingestion of 50 g glucose, the tentative glucose appearance rate was estimated using Elashoff et al. [31] meal simulation model (Figure 5.1).

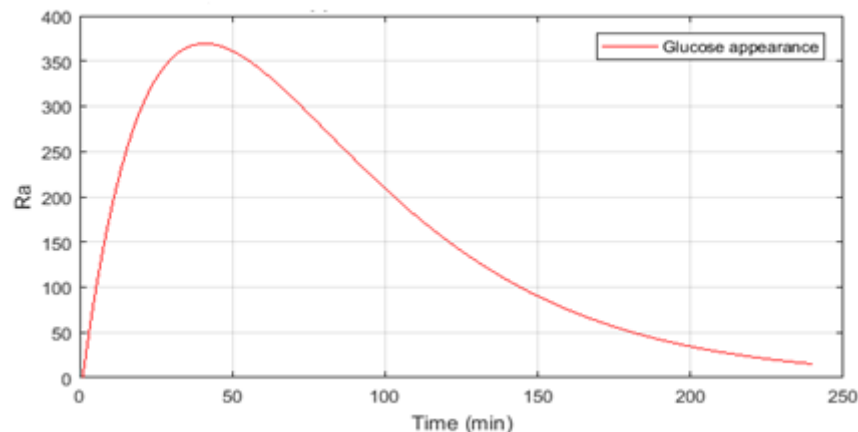


Figure 5.1: Trajectory of glucose appearance rate (mg/min) during OGTT.

Based on glucose appearance rate the model described in section 4.3 was optimized to the trajectories of plasma glucose and insulin concentration profile.

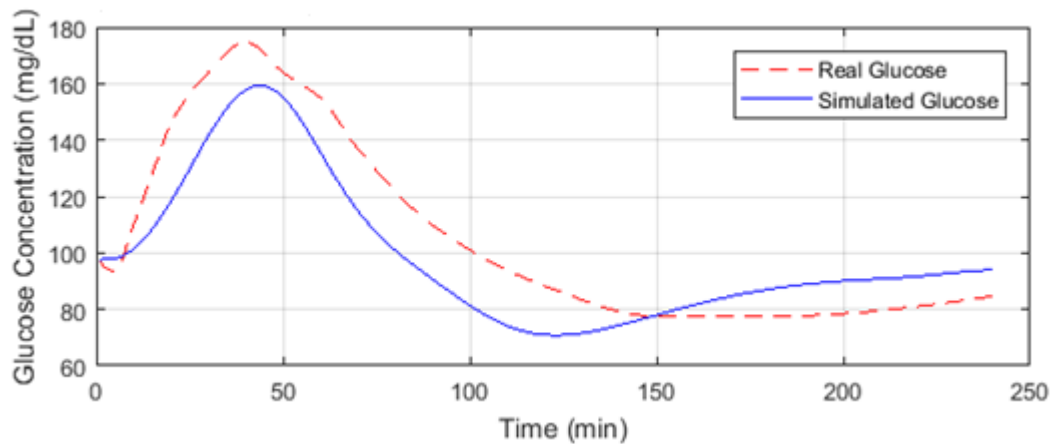


Figure 5.2: Real (dashed) vs. simulated (continuous) glucose concentration for OGTT.

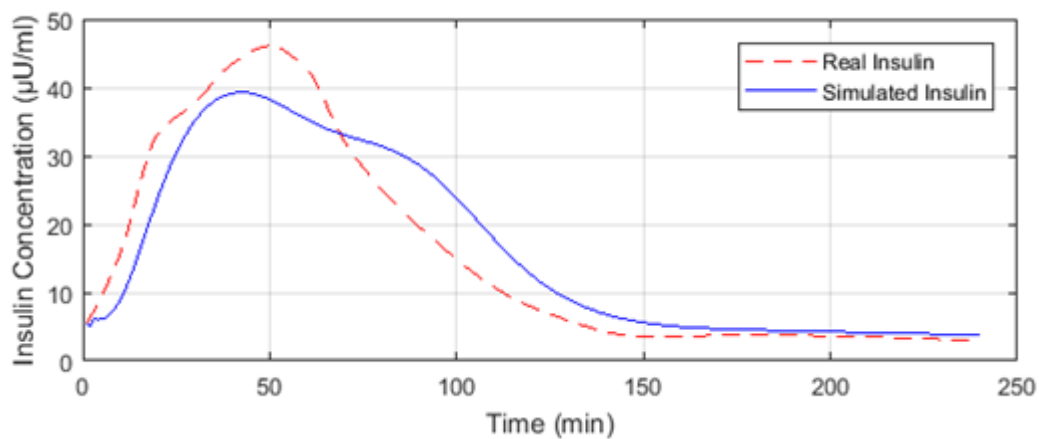


Figure 5.3: Real (dashed) vs. simulated (continuous) insulin concentration for OGTT.

Hence, from the fitting experiment, it was observed that the simulated glucose signal produces a 92% correlation with the target signal (Figure 5.2). At the same time, the simulated insulin signal produced 95% correlation with the target (Figure 5.3).

5.2 Fitting clinical exercise experiment

Figure 5.4 shows the oxygen uptake rate during a moderate level of physical activity of the selected clinical exercise experiment described in section 3.1.2. In response to this exercise intensity, a corresponding change in glucose, free fatty acid, lactate, insulin, and glucagon concentration both from model described in section 4.1 and dataset have been illustrated in Figure 5.5, Figure 5.6, Figure 5.7, Figure 5.8, and Figure 5.9 respectively.

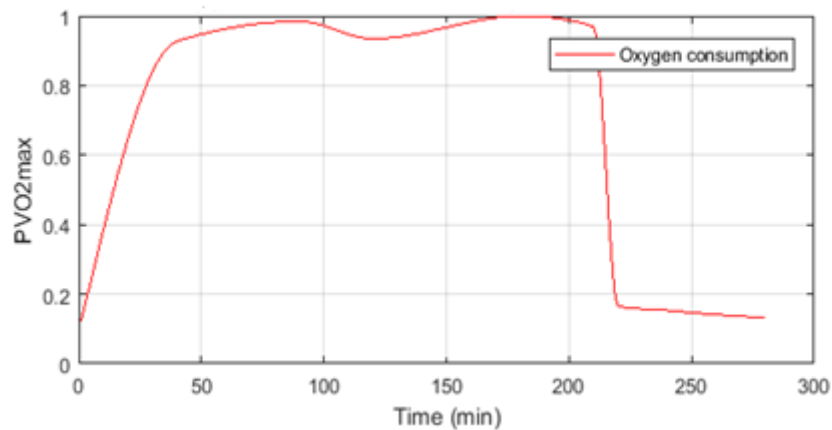


Figure 5.4: Trajectory of oxygen consumption rate during a clinical experiment of exercise.

Based on the trajectory of oxygen consumption rate during the clinical experiment of exercise the model was optimized to the trajectories of plasma glucose, free fatty acid, lactate, insulin, and glucagon concentration profile.

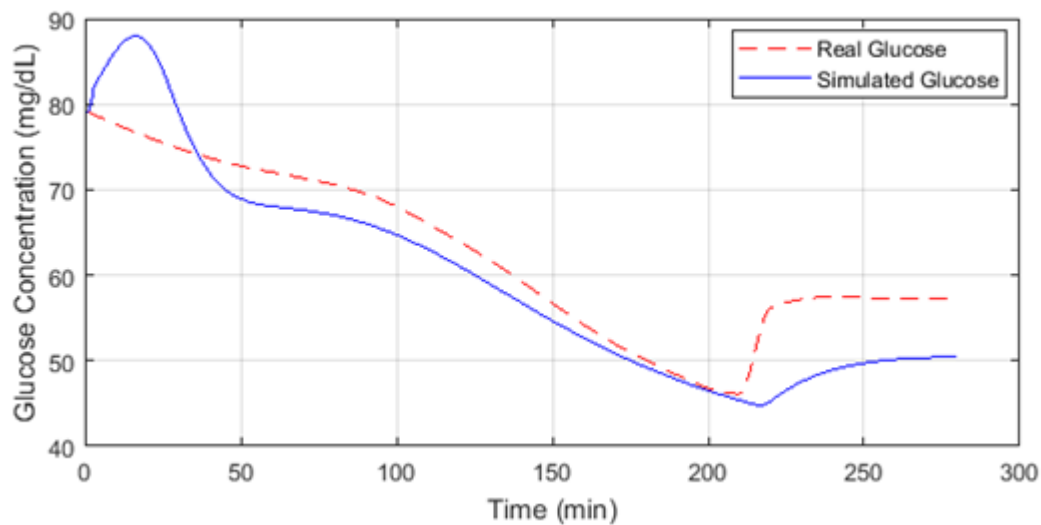


Figure 5.5: Real (dashed) vs. simulated (continuous) plasma glucose concentration for clinical exercise experiment.

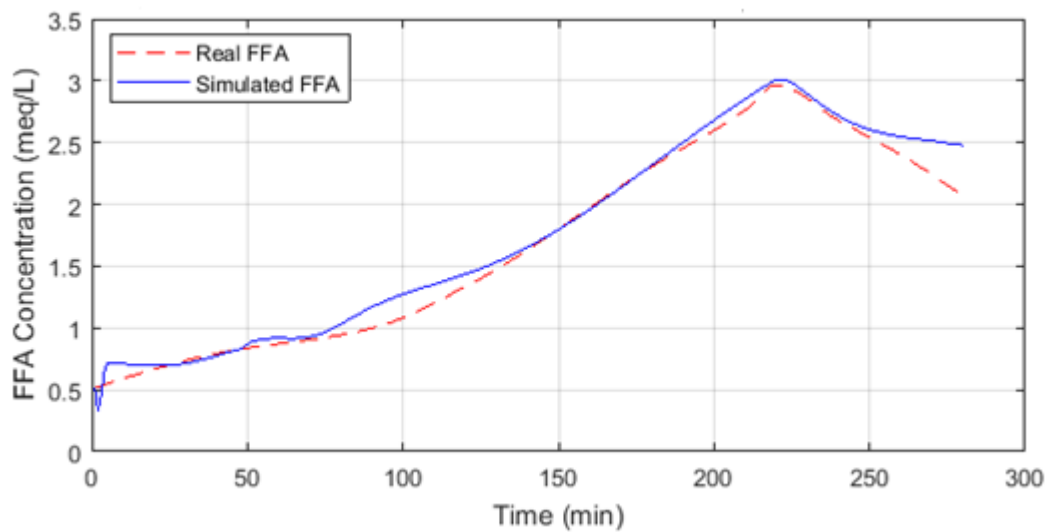


Figure 5.6: Real (dashed) vs. simulated (continuous) plasma FFA concentration for clinical exercise experiment.

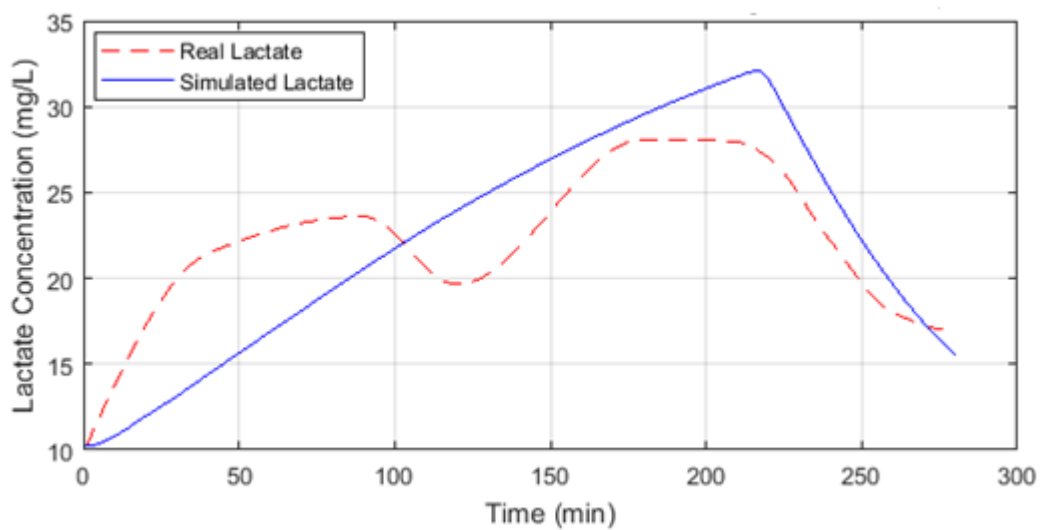


Figure 5.7: Real (dashed) vs. simulated (continuous) plasma lactate concentration for clinical exercise experiment.

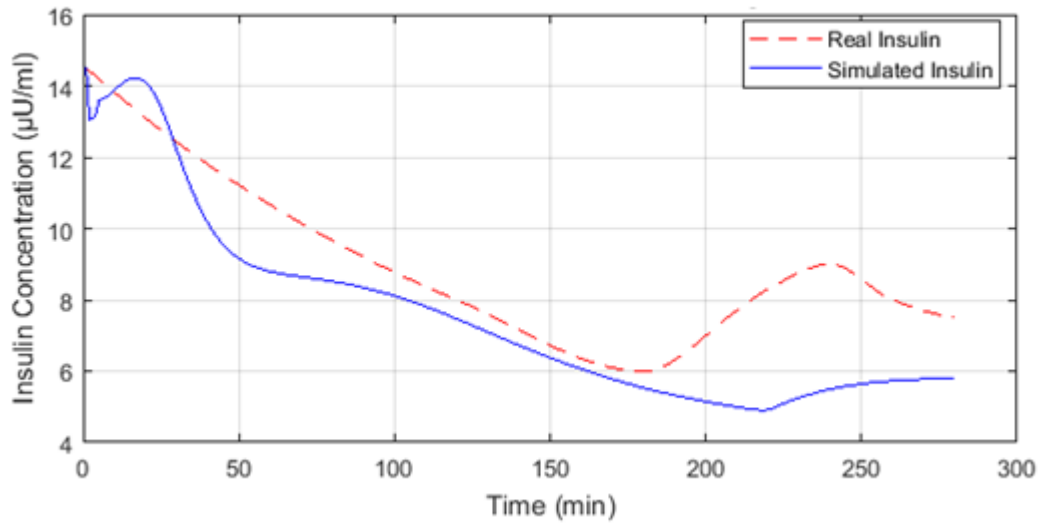


Figure 5.8: Real (dashed) vs. simulated (continuous) plasma insulin concentration for clinical exercise experiment.

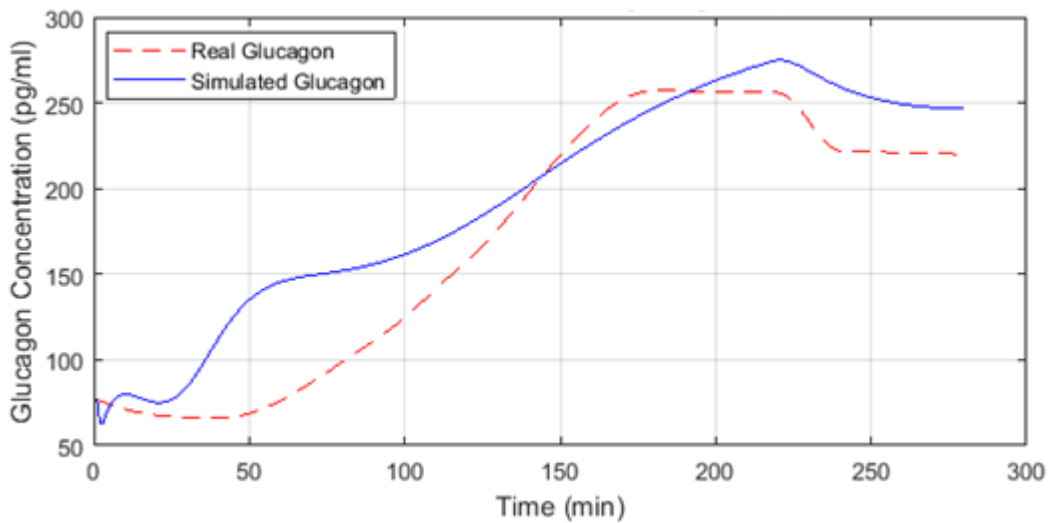


Figure 5.9: Real (dashed) vs. simulated (continuous) glucagon concentration for clinical exercise experiment.

By comparing the model response with reference concentration, it was obtained that simulated glucose, FFA, lactate, insulin, and glucagon trajectories produced a correlation of 93%, 99%, 81%, 90%, and 95% respectively with the corresponding target sequence.

5.3 Fitting continuous glucose profile

The model of section 4.3 was optimized with a daylong continuous signal of glucose appearance rate (Figure 5.10), physical activity (Figure 5.11) and corresponding plasma glucose excursion

of a type-2 diabetic patient for the assessment of model capacity.

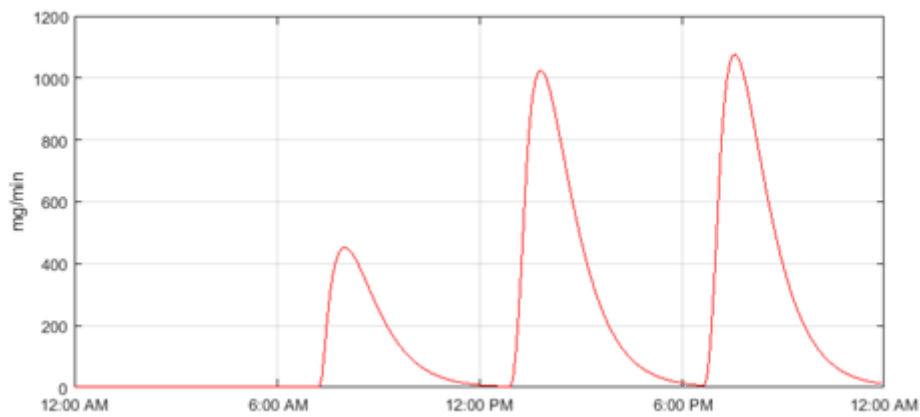


Figure 5.10: Continuous glucose appearance rate (mg/min).

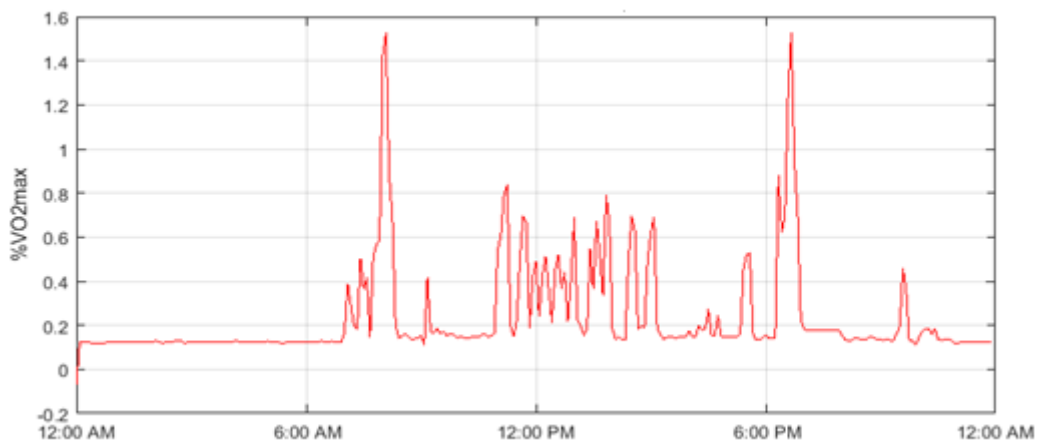


Figure 5.11: Continuous activity signal in % of maximum volume of oxygen consumption.

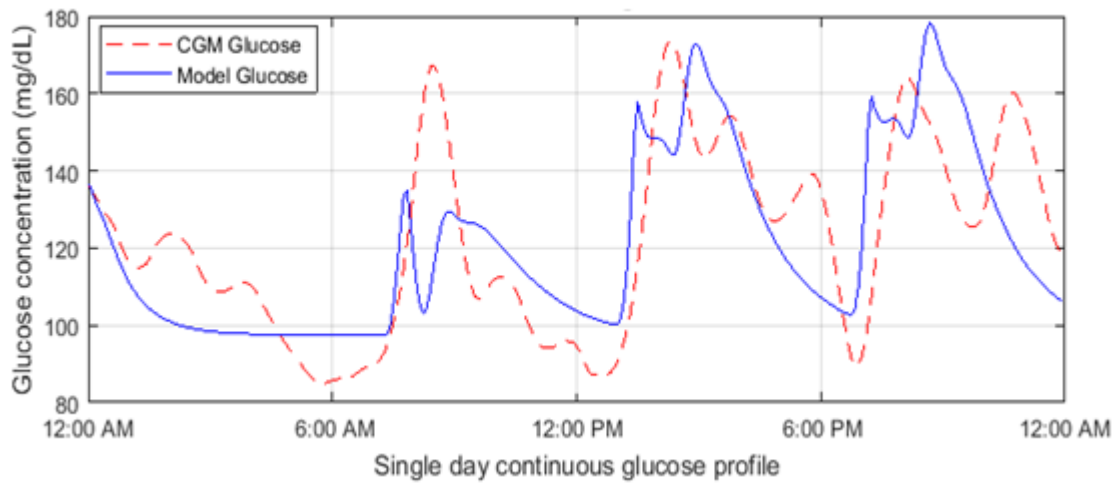


Figure 5.12: Real continuous glucose profile (dashed) vs. simulated (continuous) glucose profile for type 2 diabetes.

From the fitting result (Figure 5.12), it was found that the simulated glucose signal produces a correlation of 67% with the target concentration.

5.4 Comparison with the S.M.Ewings model

The performance of the proposed model was compared with the model of S.M. Ewings et al. [16] which was built on the assumption of no production capability of endogenous insulin. Being built on a minimal model of glucose dynamics, the S.M. Ewings model consists of six (06) differential equations and is implemented in Simulink (Figure 5.13). Continuous glucose profile along with diet, insulin, and physical activity data of a type-1 diabetic were considered for comparison.

Figure 5.15 shows the optimization result of both the proposed and S.M. Ewings et al. [16] model for single-day glucose profile of a type-1 diabetic patient against the continuous signal of carbohydrate, exogenous insulin, and activity (Figure 5.14).

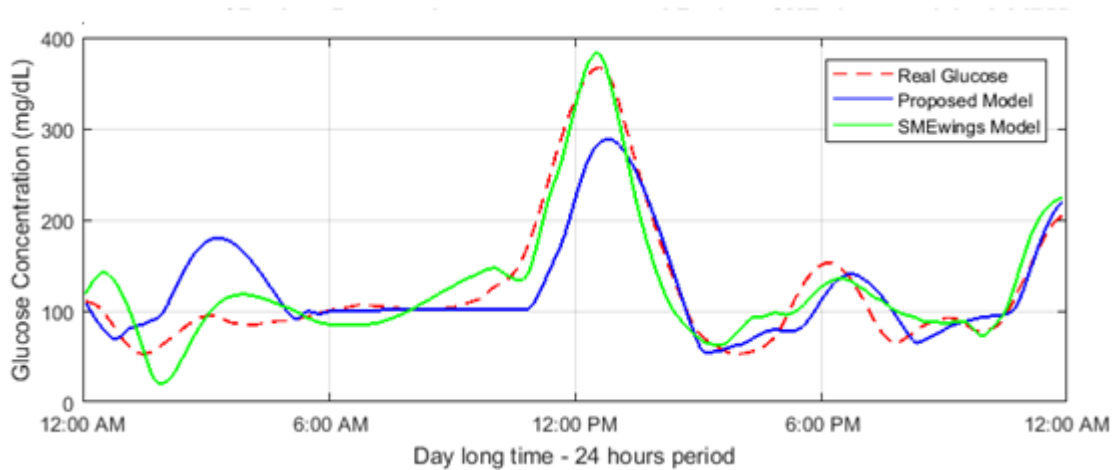


Figure 5.15: Comparison among real glucose and simulated glucose signals of a type 1 diabetic for a single day longer.

The simulated glucose profile from the proposed model produced a lower correlation coefficient (84%) compared to the SM Ewings model (95%). But the proposed model can give more metabolic insight into blood glucose dynamics by plotting the effect equations. Figure 5.16 describes the effect of plasma glucose, insulin, and glucagon on various metabolic processes in the liver. Similarly Figure 5.17, Figure 5.18, and Figure 5.19 show spectral interpretation of effect equations of muscle, pancreas, and plasma.

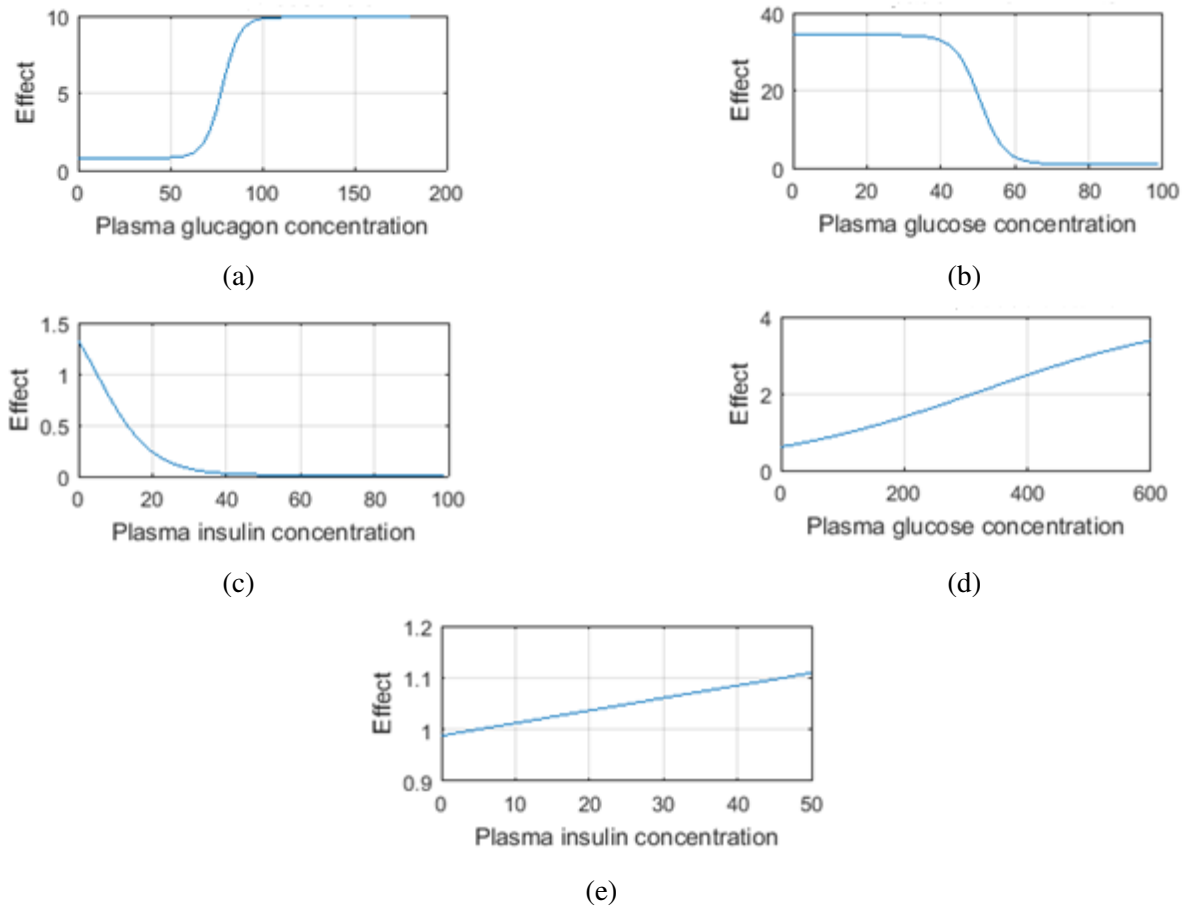


Figure 5.16: Liver compartment - Effect of glucagon (a), glucose (b) and insulin (c) on hepatic release. Effect of glucose (d) and insulin (e) on hepatic uptake.

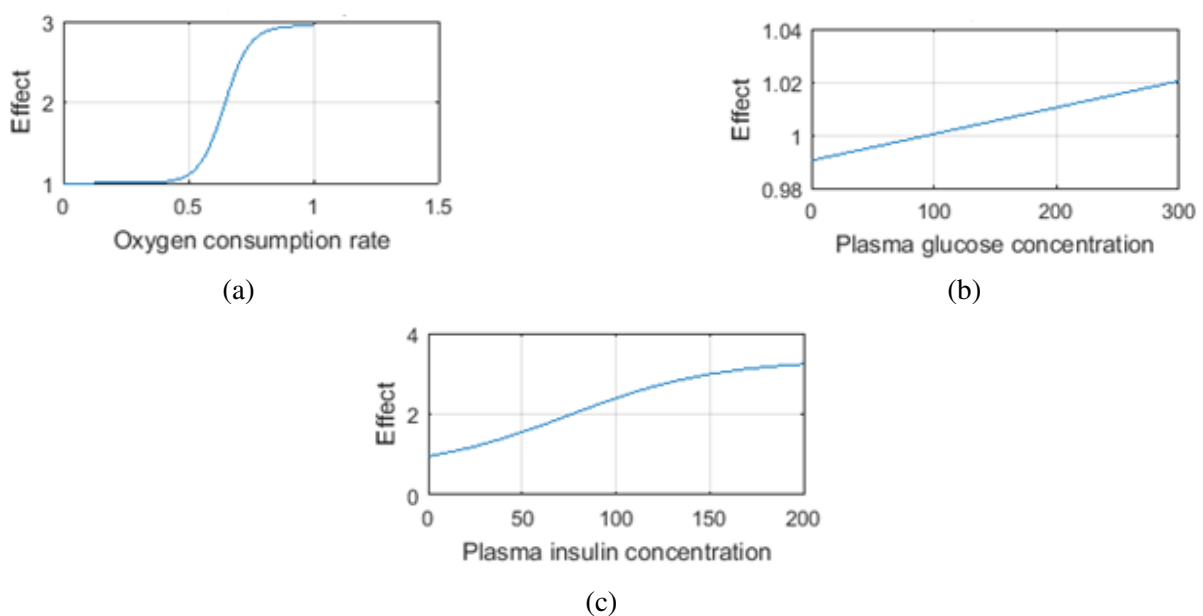


Figure 5.17: Muscle compartment - Effect of activity (a), glucose (b) and insulin (c) on muscle uptake of glucose.

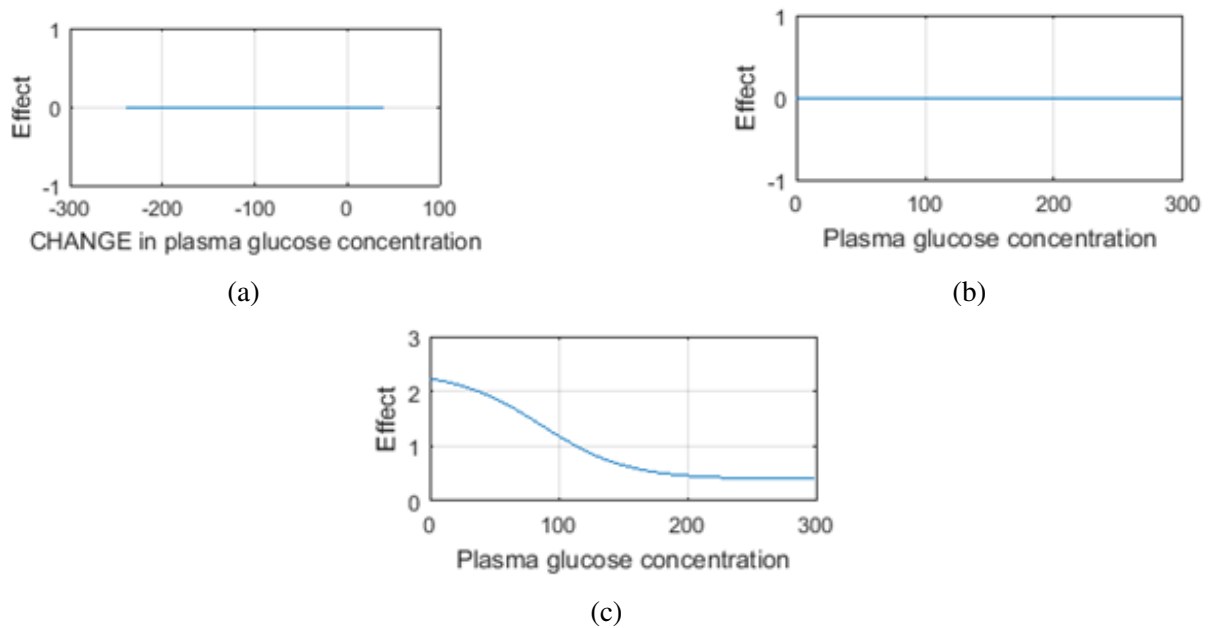


Figure 5.18: Pancreas compartment - Effect of glucose CHANGE (a) and glucose (b) on insulin release. Effect of glucose (c) on glucagon release

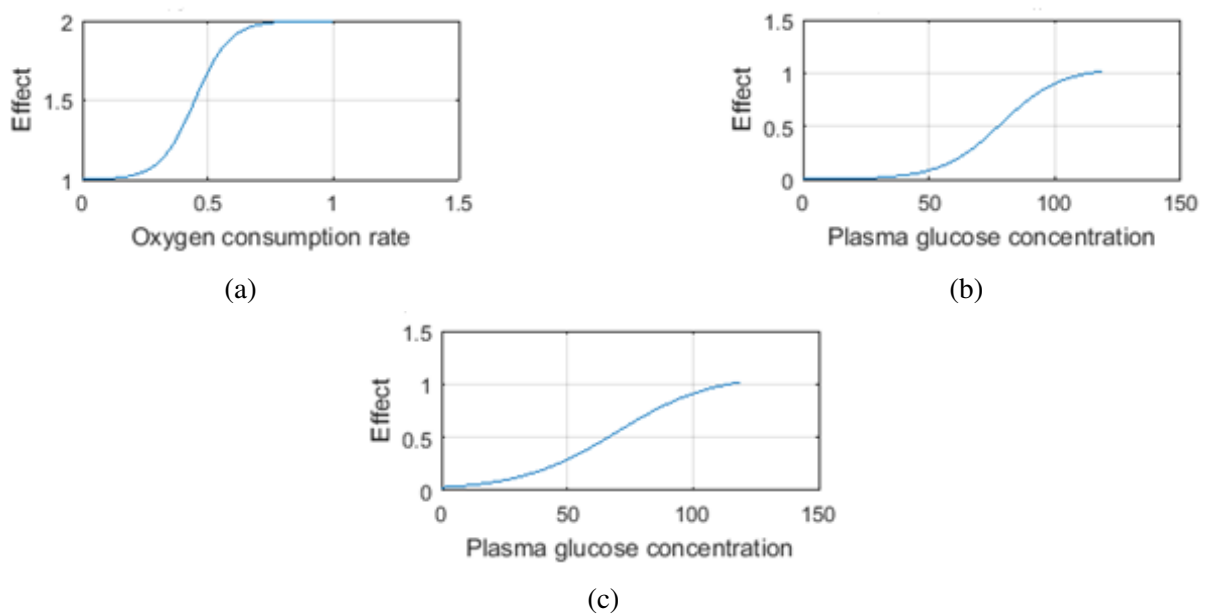


Figure 5.19: Plasma compartment - Effect of activity (a) on insulin degradation. Effect of glucose (b) on RBC uptake. Effect of glucose (c) on glucose uptake in nervous system

Parameters of the S.M.Ewings model are estimated using unconstrained nonlinear least squares optimization and mentioned in Table 5.1.

Table 5.1: List of parameters and their estimated values of S.M. Ewings model for single-day glucose profile of type 1 diabetic.

Parameters	Values	Unit	Description
p_1	2.0315	min^{-1}	Rate of insulin clearance
p_2	0.0027	min^{-1}	Rate of insulin degradation
p_3	0.0228	$ml/U/min^2$	Rate of insulin appearance
p_4	0.8	min^{-1}	Rate constant
p_5	0.8	min^{-1}	Rate constant
p_6	0.0523	mg/ml	Rate constant
p_7	0.0038	min^{-1}	Rate constant
p_8	0.0985	mg/ml	Rate constant
p_9	0.0154	min^{-1}	Rate constant
p_{10}	-0.0236	min^{-1}	Rate of Insulin independent glucose clearance
p_{11}	114	mg/dl	Basal glucose concentration

5.5 Comparison of fitting experiments

The performance of the proposed model was assessed by conducting a fitting experiment on OGTT, Clinical Exercise test, and glucose profile of free-living environment. The Dataset of those experiments were different in nature, size, and no of variables. Hence, to compare the results of all experiments correlation coefficient was used as the metric for bringing all experiments on the same platform. The result of all experiments was listed in Table 5.2.

Table 5.2: Correlation coefficient among simulated responses generated from proposed model and corresponding target in different fitting experiments.

Fitting Experiments	Input Events	Glucose	Insulin	Glucagon	FFA	Lactate
OGTT	Oral glucose	0.92	0.95	–	–	–
Exercise	O2 consumption	0.94	0.90	0.95	0.99	0.81
Type-2 glucose profile	Diet, Activity	0.67	–	–	–	–
Type-1 glucose profile	Diet, Insulin, Activity	0.84	–	–	–	–
Average Correlation		0.84 ± 0.12	–	–	–	–

5.6 Effect of correlation variation in ML models

To experiment with the effect of correlation variation on the performance of the ML-based predictive model, two FFNN models of the same architecture (section 3.2) are built on two separate datasets with various prediction horizons (PH). Along with the same CGM and diet information, one dataset contains real activity signal (Figure 3.6) and another one contains the synthetic activity signal (Figure 3.8). The synthetic activity has a greater negative correlation of about 11.24% than 0.08% of real activity signal with the CGM data as displayed in scatter plot in (Figure 3.9 and Figure 3.7). Both models are trained over 15 days of the dataset and tested on the rest. Performances of two models are evaluated using RMSE and Clarke Error Grid Analysis and described in Table 5.3.

Table 5.3: Accuracy comparison of two FFNN of same architecture trained on two different datasets (real & synthetic) of different correlation over various PH for 05 full Days.

PH	Model	RMSE	Zone-A	Zone-B	Zone-C	Zone-D	Zone-E
2hr	Real	27.24 ± 1.47	57 ± 0.08%	42 ± 0.09%	0%	1 ± 0.01%	0%
	Synthetic	18.15 ± 2.76	82 ± 0.07%	17 ± 0.07%	0%	1 ± 0.01%	0%
3hr	Real	27.23 ± 1.64	56 ± 0.08%	43 ± 0.07%	0%	1 ± 0.01%	0%
	Synthetic	23.01 ± 2.69	69 ± 0.09%	31 ± 0.09%	0%	1 ± 0.01%	0%
4hr	Real	28.49 ± 2.22	52 ± 0.08%	47 ± 0.07%	0%	1 ± 0.01%	0%
	Synthetic	26.46 ± 3.79	59 ± 0.12%	40 ± 0.11%	0%	1 ± 0.01%	0%
5hr	Real	34.44 ± 5.45	54 ± 0.06%	45 ± 0.05%	0%	1 ± 0.01%	0%
	Synthetic	30.01 ± 6.46	55 ± 0.11%	45 ± 0.10%	0%	1 ± 0.01%	0%
6hr	Real	30.65 ± 2.78	53 ± 0.06%	46 ± 0.05%	0%	1 ± 0.01%	0%
	Synthetic	30.35 ± 3.66	59 ± 0.07%	41 ± 0.07%	0%	1 ± 0.01%	0%

* Data Format = Mean ± Standard Deviation; Real → Lower Correlation; Synthetic → Higher Correlation;

From Table 5.3, it is observed that most of forecasted values are located in the clinically acceptable region (Zone-A and Zone-B) in the error grid analysis for both models. Very few forecasted points are located in Zone-D. It is also obvious that the RMSE of forecasted glucose trajectories increases for both models as the length of the prediction horizon increases. Though both FFNN models are of the same architecture, the performance of the FFNN model trained on the dataset with synthetic activity is more accurate in comparison to the performance of the FFNN model trained on the dataset with real activity signal. The obvious reason behind this is the higher correlation coefficient of activity signal with continuous glucose data for synthetic activity.

5.7 OR on ML-based forecasting

For conducting OR on the response obtained from the FFNN model of section 3.2, the forecasted glucose trajectories with three (03) hours of prediction horizon (PH) are considered. Each day of the forecasted profile is segmented into eight (08) intervals. Then the forecasted response of each interval is targeted for optimization in the physiological model (section 4.3) along with the corresponding carbohydrate on board, continuous activity signal, and estimated personalized constraints illustrated in section 4.3.4. The optimization problem is solved using sequential quadratic programming. The fitted response (OR response) of glucose concentration is compared with the corresponding actual glucose concentration sequence using RMSE and Clarke Error Grid Analysis [19] as the performance metric. By estimating the RMSE and Error Grid Analysis of ML-based forecasting with corresponding actual CGM profile, a comparison between ML-based statistical and OR-based physiological response is created (Table 5.4) for performance analysis of the proposed solution of this research. Figure 5.20 shows the OR response for corresponding ML forecasting for the time interval of 12:00 pm – 3:00 pm of day-1 from the testing part of the dataset.

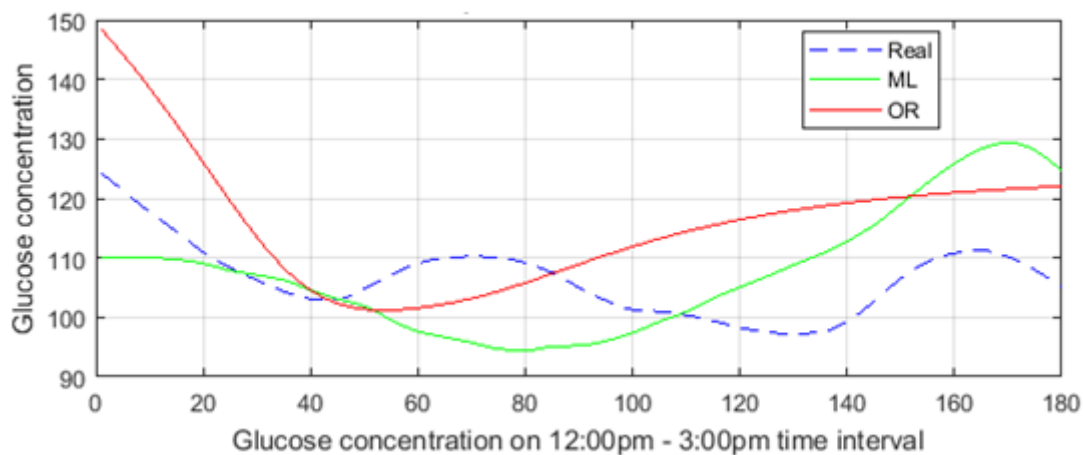


Figure 5.20: Producing OR response by optimizing physiological model.

As a result of producing OR response by optimizing physiological model on ML response, following illustration of metabolic properties is obtained according to the architecture of the physiological model. Figure 5.21 describes the effect of plasma glucose, insulin, and glucagon on various metabolic processes in the liver. Similarly Figure 5.22, Figure 5.23, and Figure 5.24 show spectral interpretation of effect equations of muscle, pancreas, and plasma.

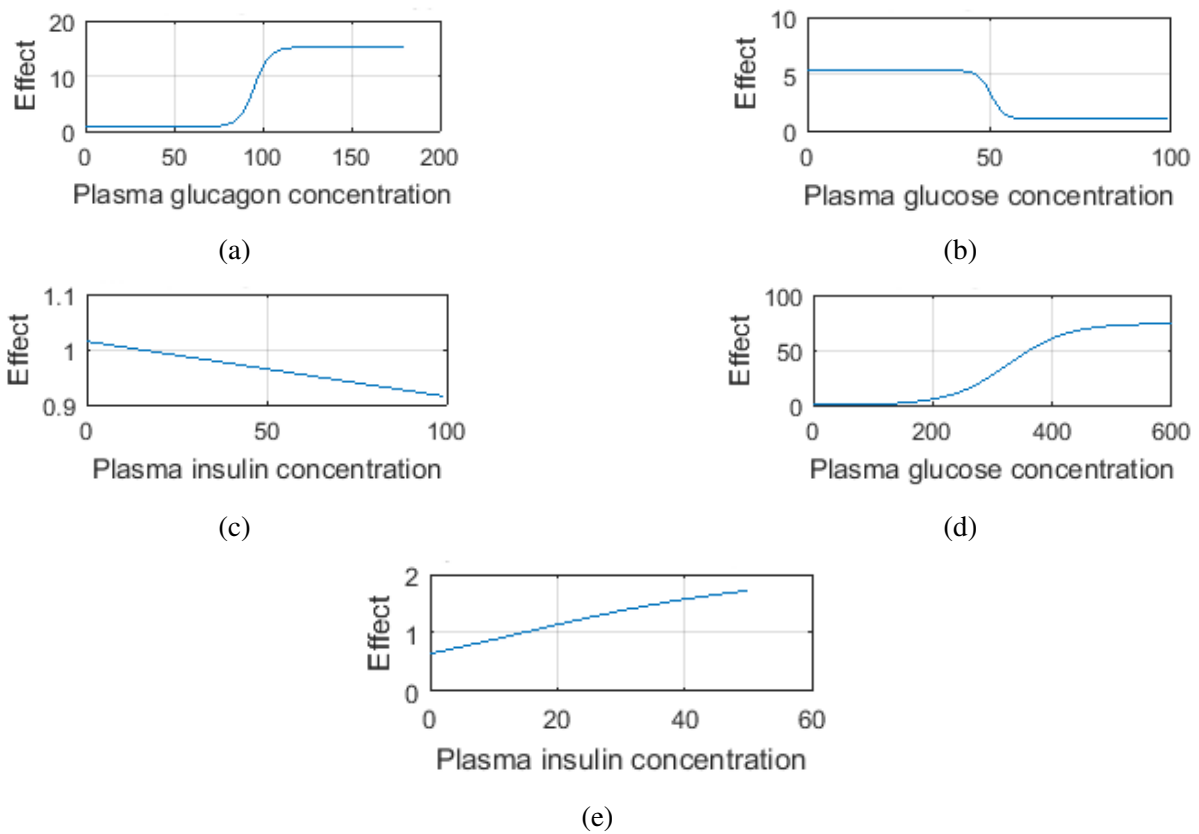


Figure 5.21: Liver compartment - Effect of glucagon (a), glucose (b) and insulin (c) on hepatic release. Effect of glucose (d) and insulin (e) on hepatic uptake.

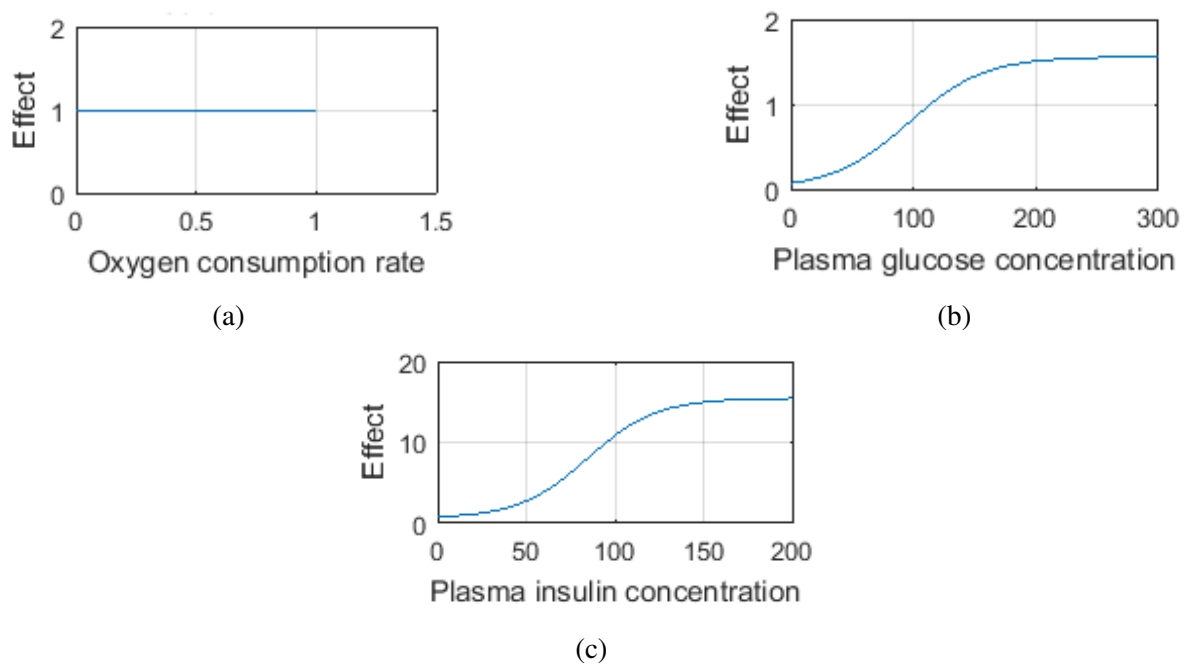


Figure 5.22: Muscle compartment - Effect of activity (a), glucose (b) and insulin (c) on muscle uptake of glucose.

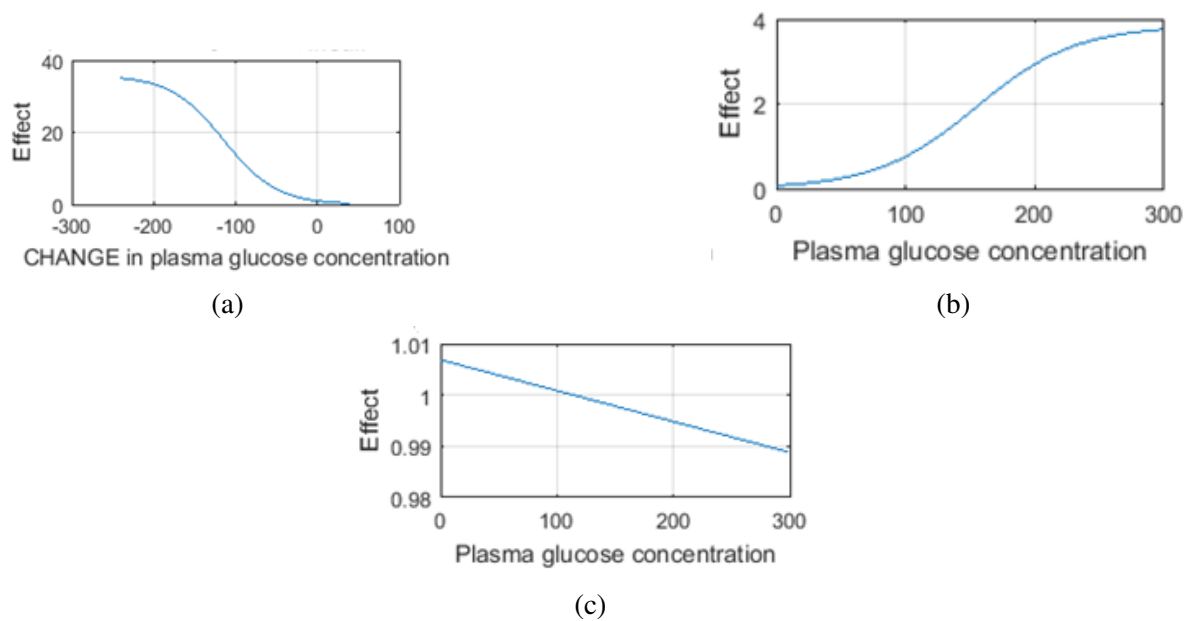


Figure 5.23: Pancreas compartment - Effect of glucose CHANGE (a) and glucose (b) on insulin release. Effect of glucose (c) on glucagon release

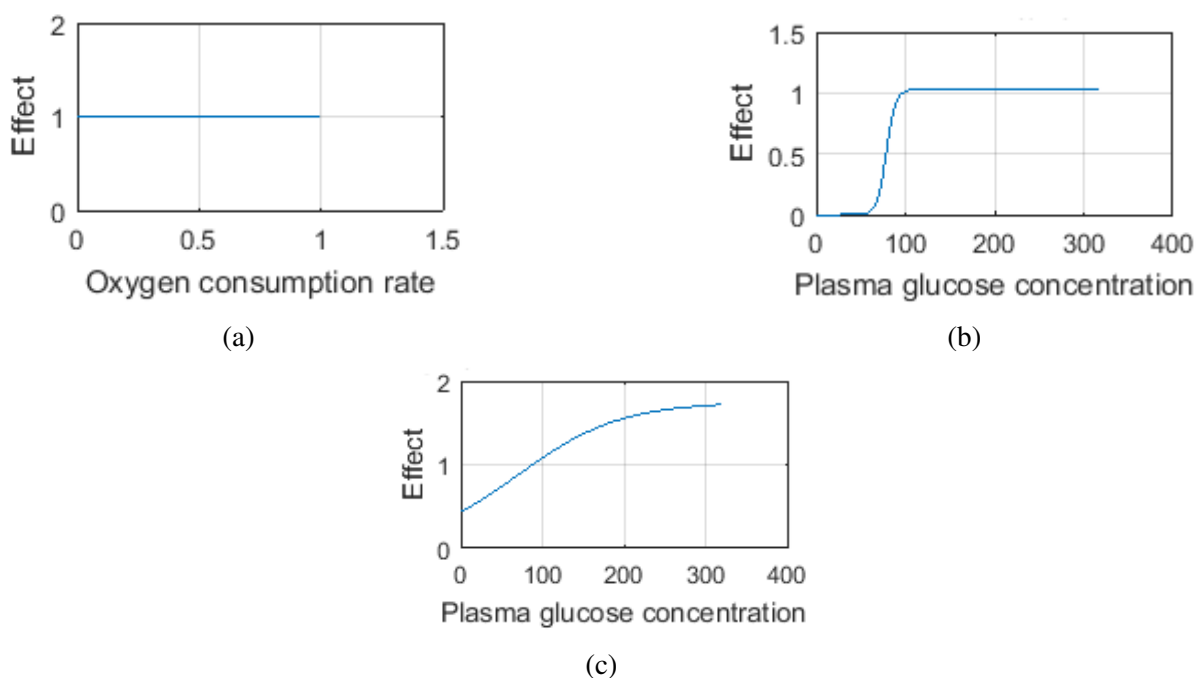


Figure 5.24: Plasma compartment - Effect of activity (a) on insulin degradation. Effect of glucose (b) on RBC uptake. Effect of glucose (c) on glucose uptake in nervous system

Conducting similar operations over forecasting of five (05) full days, this type of spectrum with different results is achieved. Table 5.4 shows the average comparative results of ML and OR models on eight (08) different time intervals in mean \pm standard deviation format over 05 full days. Figure 5.25 shows the scatter plot of Clarke Error Grid Analysis of both OR response and

ML response.

Table 5.4: Comparison among ML response vs. target & OR response vs. target over 03 hours PH for 05 full days.

Interval	Model	RMSE	Zone-A	Zone-B	Zone-C	Zone-D	Zone-E
12:00 am	ML	20.98 ± 11.93	65.11 ± 37.80	34.89 ± 37.80	0.00	0.00	0.00
-3:00 am	OR	20.64 ± 14.13	65.67 ± 44.75	34.33 ± 44.75	0.00	0.00	0.00
3:00 am	ML	10.60 ± 4.01	91.22 ± 12.29	8.78 ± 12.29	0.00	0.00	0.00
-6:00 am	OR	13.63 ± 7.72	74.33 ± 30.25	25.67 ± 30.25	0.00	0.00	0.00
6:00 am	ML	17.33 ± 7.63	78.22 ± 30.19	20.89 ± 29.29	0.00	0.89 ± 1.99	0.00
-9:00 am	OR	19.72 ± 6.98	70.67 ± 24.04	28.44 ± 23.41	0.00	0.89 ± 1.99	0.00
9:00 am	ML	24.44 ± 6.46	56.00 ± 28.58	39.89 ± 20.43	0.00	4.11 ± 9.19	0.00
-12:00 pm	OR	29.98 ± 5.78	27.44 ± 29.23	68.44 ± 25.63	0.00	4.11 ± 9.19	0.00
12:00 pm	ML	22.39 ± 16.09	64.00 ± 39.82	36.00 ± 39.82	0.00	0.00	0.00
-3:00 pm	OR	22.89 ± 11.34	57.33 ± 42.32	42.67 ± 42.32	0.00	0.00	0.00
3:00 pm	ML	25.61 ± 7.15	57.00 ± 21.37	43.00 ± 21.37	0.00	0.00	0.00
-6:00 pm	OR	22.39 ± 8.94	61.22 ± 30.32	38.78 ± 30.32	0.00	0.00	0.00
6:00 pm	ML	23.89 ± 8.40	53.22 ± 31.26	46.78 ± 31.26	0.00	0.00	0.00
-9:00 pm	OR	31.75 ± 10.41	49.78 ± 19.53	50.22 ± 19.53	0.00	0.00	0.00
9:00 pm	ML	29.47 ± 7.61	53.22 ± 12.91	45.78 ± 11.48	0.00	1.00 ± 2.24	0.00
-12:00 am	OR	25.93 ± 12.75	64.44 ± 21.70	33.44 ± 21.05	0.00	2.11 ± 4.72	0.00

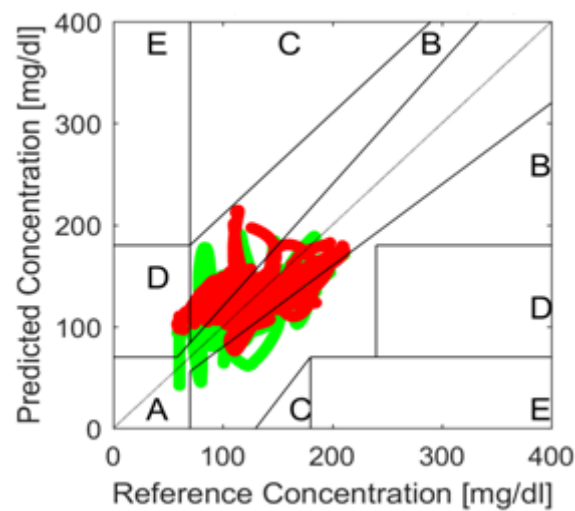


Figure 5.25: Clarke Error grid analysis of OR Response (Red) and ML Response (Green) for five (05) days.

5.8 Discussion

The main theme of this thesis is solving a problem by filling a research gap in modeling diabetes. A constraint-based blood glucose regulation model as part of IPM (section 1.4) of virtual diabetic

patient is built and validated through fitting experiments. Then the model is used in operation research for producing physiological explanations from ML-based forecasting. Hence, the discussion is segmented into two sections focusing on a separate issue.

5.8.1 Building and validation of the physiological model

In OGTT, glucose regulation was interrupted through external glucose ingestion. In the clinical exercise experiment glucose dynamics was stressed to negotiate the requirement of muscle tissue. In continuous glucose variation, glucose regulation was gone through consecutive events of oral ingestion, activity, and subcutaneous injection of insulin. Since the proposed model generated an average correlation coefficient of 0.84 ± 0.12 on simulated responses with the target (Table 5.2) and responses are analogous to the physiology described in section 3.3, the validity of the model can be claimed.

A noticeable issue is that fitting result is not as much satisfactory for day long continuous glucose profile as fitting with OGTT and Exercise experiment. The main cause of this phenomenon is the lack of ability of mathematical model for capturing temporal variation of metabolic behavior. The time length of OGTT and Exercise experiment is less than five (05) hours. The change in temporal variation of metabolic behavior is not significant for five (05) hours. But due to circadian rhythm, there is a diurnal variation in insulin sensitivity in the human body. As a result day long variation in metabolic behavior is significant.

The utility of the proposed model comes out while comparing it with the reference model. The proposed model can produce more metabolic insights (Figure 5.16 - Figure 5.19) than that of the reference model (Table 5.1). Parameters of the reference model (Table 5.1) are static rate constant whereas effect equations of the proposed model illustrate nonlinear saturation relation of plasma variables on basal rate of metabolic process. The reference model considers only hepatic release and peripheral uptake process to integrate the impact of activity in glucose dynamics. In absence of activity, the reference model can give information only on glucose clearance due to insulin and independent of insulin. But the proposed model can describe the glucose distribution on more metabolic processes with the nonlinear effect of plasma variables.

Once the model is optimized all effect equations collectively provide a metabolic spectrum. By fitting the model on a continuous glucose profile on a segment basis, it is possible to monitor deviation of trajectories of effects of plasma variables on metabolic rates. This spectrum may help to get more information of temporal variation on the metabolic condition of a patient. Similar procedure for the assessment of diabetic abnormalities in type-2 diabetic patient is shown by O. Vahidi [8] in his PhD dissertation. In Figure 5.26 – Figure 5.28 mathematical

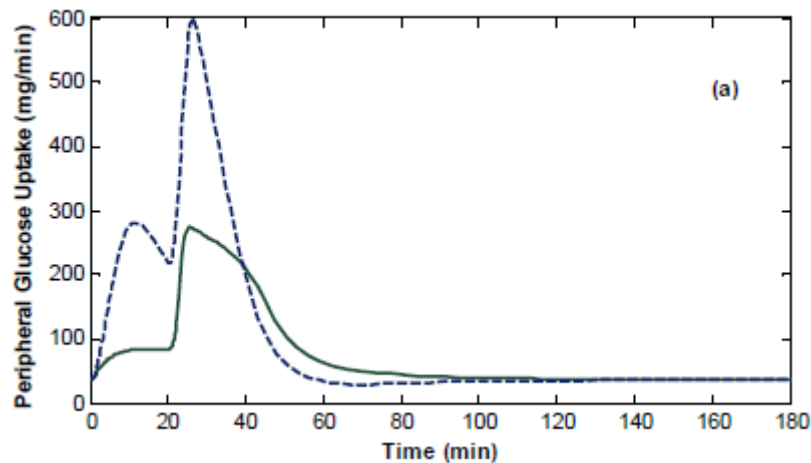


Figure 5.26: Peripheral glucose uptake rate during the FSIGT test: type II diabetic patients (-) and healthy subjects (-).

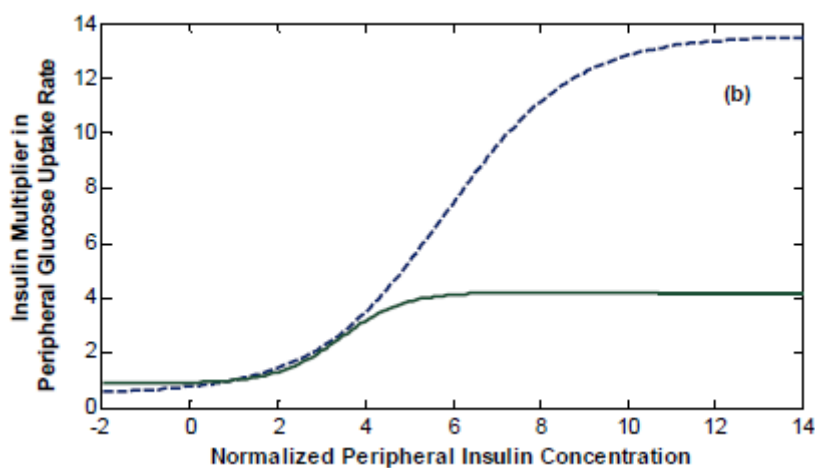


Figure 5.27: Insulin multipliers in peripheral glucose uptake rate versus normalized plasma insulin concentrations, type II diabetic patients (-) and healthy subjects (-).

evaluation of peripheral glucose uptake rate in a healthy and diabetic patient in a frequently sampled insulin-modified intravenous glucose tolerance test (FSIGT) taken from [8] is shown. It is obvious from Figure 5.27 and Figure 5.28 that the effect of plasma glucose and insulin concentration on peripheral glucose uptake rate is different in diabetic and healthy subjects. As a consequence, glucose uptake rate is also different.

The proposed model can be used as a research tool for preparing in silico environments for modeling diabetes. Producing a glucose profile with many hypoglycemic/ hyperglycemic events of a type-2/type-1 diabetic is unethical and dangerous. Hence producing a synthetic dataset, the proposed model of this paper can be leveraged. Most potential usage of the proposed constraint-based structure is the OR on ML-based forecasting intending to obtain physiological

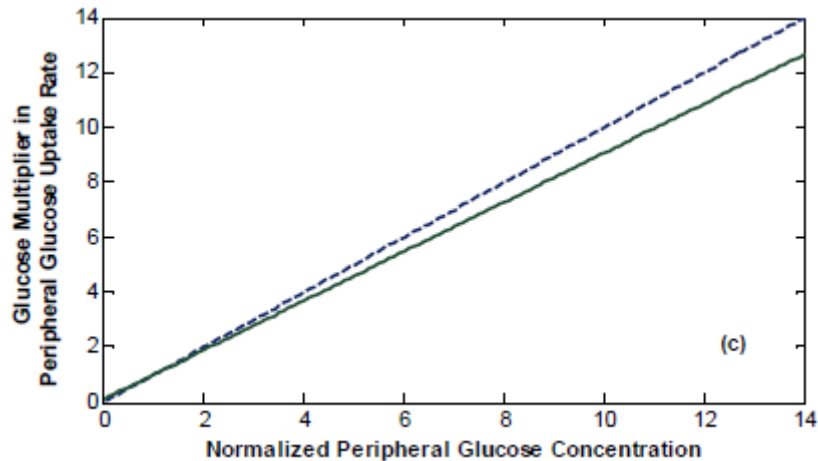


Figure 5.28: Glucose multipliers in peripheral glucose uptake rate versus normalized plasma glucose concentrations, type II diabetic patients (-) and healthy subjects (-).

interpretation.

There are some unavoidable issues still unexplored. Specifically, induction of insulin sensitivity due to exercise is represented with a sigmoid function of effort which is the integrated product of exercise intensity and duration. But quantitative analysis of induction and reduction of insulin sensitivity does not appear in the experiments. Insulin sensitivity is a very important metabolic factor in the blood glucose regulation of a diabetic patient which has a natural diurnal variation and is also greatly influenced due to physical activity. It has a very crucial role in estimating the appropriate amount of insulin dose at a particular time of the day for an insulin-dependent diabetic patient. The mathematical architecture of the proposed model is constructed without considering the age, gender, injury, and mental condition of diabetic patients. Even no particular function has been incorporated for capturing diurnal variation of insulin sensitivity.

5.8.2 OR on ML for physiological interpretation

By performing OR it is possible to transform the data-driven forecasted profile of glucose concentration into the metabolic signal (Figure 5.20) which shows physiological interpretation through producing the relation between plasma variables and metabolic process rate according to Table 4.1 (Figure 5.21 – Figure 5.24). If a forecasted sequence of glucose concentration data is undergone for constrained nonlinear optimization, the internal parameters of the physiological model will be optimized in such a way that sigmoid functions of the model represent the trajectory of the effects for the change of concentration of plasma variables. Once the model is optimized for an interval, all hyperbolic tangent functions collectively provide a metabolic spectrum of the blood glucose regulation for that specific period. That spectrum usually interprets how the concentration of plasma glucose, insulin, glucagon, and exercise influence the metabolic

processes in various organs/tissues. The personalized constraints are the maximum magnitude of effect trajectories estimated on the real profile for a specific person. Optimization under these constraints gives reliability and personalization to produced OR response.

By assessing the tabular result (Table 5.4) it is obvious that RMSE was usually increased while performing OR for most of the cases. But both responses were located in the clinically acceptable zone (zone-A and Zone-B) for maximum time intervals in Clarke Error Grid Analysis (Figure 5.25). The reason behind the increment of RMSE in OR response is the compliance of four (04) types of constraints during optimization. Besides this, very limited factors responsible for the glycemic excursion were incorporated in the physiological model. Hence, it was harder to produce the real trajectories in OR response by complying with all the constraints taken under consideration.

In the time interval of 9:00 am -12:00 pm, some clinically dangerous responses were produced in both ML and OR responses. That means both ML and OR models suffered to produce real plasma glucose responses in that time interval. By reviewing the glucose profile of that interval it appears that there may not have sufficient event information regarding diet events or correct recording of glycemic variation. From Table 5.4 it is also obvious that the RMSE result is much better at 9:00 pm – 9:00 am than that of 9:00 am to 9:00 pm. The cause behind this improvement is that the first 12 hours (9:00 pm – 9:00 am) of the day is night time and glucose regulation is less disturbed by external events compared to the next 12 hours (9:00 am – 9:00 pm).

The result obtained in error grid analysis produces inspiration for moving forward in applying OR on ML-based forecasting for taking the therapeutic decision in diabetes management. Once the OR model is optimized for forecasted glucose trajectories it is possible to simulate the model with various types of external disturbances to observe the response for deciding the magnitude of action for diabetes management. To the best of the author's knowledge, hybridization of OR and ML has never been done for building a predictive model of diabetes in the literature. This study extends the knowledge in the area of diabetes modeling by illustrating a way of applying ML-based forecasting in taking a therapeutic decision in diabetes management.

Chapter 6

Conclusion

6.1 Summary and findings

With a vision of building a reliable support system for taking appropriate action in diabetes management, this thesis is subjected to build a hybrid model combining the advantages of both ML and physiological models. At the beginning of the research, a rigorous study on ML-based predictive models is done to find the strength and limitations of the models in the field of modeling diabetes. As a result of the study, it is found that robustness analysis and physiological interpretation are the two missing features of ML models. These features are essential for successful application of the predictive models in glucose control. Data-driven experimentation is done to observe the result of correlation variation in ML-based forecasting.

OR on the forecasted trajectory is defined as the solution for enhancing the capability of an ML-based predictive model. A constraint-based mathematical model of glucose dynamics with appropriate external stimulus is required for OR and that type of model is found unavailable in the literature. Hence, a constraint-based glucose regulation model consisting of external glucose, insulin and exercise stimulus and capable of describing desired diabetic behavior is built and validated for implementing the proposed solution. Since the glucose regulation model is integrated with existing physiological models of external stimulus from literature, this composition is defined as integrated physiological model (IPM) of virtual diabetic patient. Validation of the IPM is performed using OGTT, clinical exercise experiment, and CGM dataset. It is found that the constraint-based glucose dynamics model produces an average correlation of 0.84 ± 0.12 on fitting experiments. Besides this, the proposed model can give more metabolic insight than the reference model of the literature.

An FFNN trained on CGM profile, diet and, activity information of a type-2 diabetic patient is used to produce ML-based forecasting. This forecasting is optimized with the physiological

model through constrained nonlinear optimization. The effect equations of the physiological model produce the metabolic spectrum of plasma variables and their effect on metabolic process rate. In the quantitative analysis of the performance of OR on ML, it is found that RMSE is increased in OR response compare to ML response. But a satisfactory result is found in Clarke Error Grid Analysis.

6.2 Recommendation for future works

Since the proposed IPM is validated against only few clinical trials, it is required to take more attempts to assess the performance of the model in more critical clinical experiments. Particularly the induction of insulin sensitivity due to physical exercise has not been assessed due to the unavailability of experimental data. Similarly, the long-term impact of exercise on insulin sensitivity is also needed to be explored in the future.

Since the proposed glucose regulation model of IPM produces the effect of concentration of substrates on the rate of various metabolic processes, it is possible to derive more information regarding the medical condition of a diabetic patient by assessing the rate of change of metabolism obtained in mathematical optimization results. So, the utility of the model can be further enhanced based on this approach.

By assessing the result of performing OR on ML-based forecasting it is obvious that though the OR responses are located in the clinically acceptable region in error grid analysis, the RMSE value is not up to the mark particularly in the time interval that is more interrupted with diet, insulin, and activity events. Hence this issue is needed to be explored for further improvement. There is another issue to be focused on. To produce physiological interpretation from ML-based forecasted trajectories, the proposed model has been only evaluated with data from one diabetic patient. This evaluation process is required to be performed on a more diabetic patient of different age, gender, mental, and health condition to validate the solution from a larger perspective. Hence, this can be another avenue to go forward in the research field.

References

- [1] International Diabetes Federation, “Advocacy guide to the IDF diabetes atlas 9th edition 2019.”
- [2] American Diabetes Association, “Diagnosis and classification of diabetes mellitus,” vol. 31, pp. 1935–5548, 2008.
- [3] A. Z. Woldaregay *et al.*, “Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes,” *Artificial intelligence in medicine*, vol. 98, p. 109–134, 2019.
- [4] P. Gyuk, I. Vassanyi and I. Kosa, “Blood glucose level prediction for diabetics based on nutrition and insulin administration logs using personalized mathematical models,” *Journal of Healthcare Engineering*, vol. 2019, no. 8605206, p. 12 pages, 2019.
- [5] C. Liu *et al.*, “Long-term glucose forecasting using a physiological model and deconvolution of the continuous glucose monitoring signal,” *Sensors (Basel, Switzerland)*, vol. 19(19), no. 4338, 2019.
- [6] A. N. Sveshnikova, M. A. Panteleev, A. V. Dreval , T. P. Shestakova, O. S. Medvedev, and O. A. Dreval, “Theoretical evaluation of the parameters of glucose metabolism on the basis of continuous glycemia monitoring data using mathematical modeling,” *BIOPHYSICS*, vol. 62, no. 5, p. 842–847, 2017.
- [7] Y. Zhang, T. A. Holt, and N. Khovanova, “A data driven nonlinear stochastic model for blood glucose dynamics,” *Computer Methods and Programs in Biomedicine*, vol. 125, pp. 18–25, 2016.
- [8] O. Vahidi, *Dynamic Modeling of Glucose Metabolism for the Assessment of Type II Diabetes Mellitus*. PhD thesis, University of British Columbia, 2013.
- [9] R. Visentin *et al.*, “The UVA/Padova type 1 diabetes simulator goes from single meal to single day,” *Journal of diabetes science and technology*, vol. 12, no. 2, p. 273–281, 2018.
- [10] R. Hovorka *et al.*, “Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes,” *Physiological measurement*, vol. 25, no. 4, p. 905–920, 2004.

- [11] J. T. Sorensen, *A Physiologic Model of Glucose Metabolism in Man and its use to Design and Assess Improved Insulin Therapies for Diabetes*. PhD thesis, Massachusetts Institute of Technology, 1985.
- [12] P. G. Fabietti, V. Canonico, M. O. Federici, M. M. Benedetti, and E. Sarti, "Control oriented model of insulin and glucose dynamics in type 1 diabetics," *Medical & biological engineering & computing*, vol. 44, p. 69–78, 2006.
- [13] A. Roy and R. S. Parker, "Dynamic modeling of exercise effects on plasma glucose and insulin levels," *Journal of diabetes science and technology*, vol. 1, no. 3, p. 338–347, 2007.
- [14] M. D. Breton, "Physical activity-the major unaccounted impediment to closed loop control," *Journal of diabetes science and technology*, vol. 2, no. 1, p. 169–174, 2008.
- [15] C. D. Man, M. D. Breton, and C. Cobelli, "Physical activity into the meal glucose-insulin model of type 1 diabetes: in silico studies," *Journal of diabetes science and technology*, vol. 3, no. 1, pp. 56–67, 2009.
- [16] S. M. Ewings, S. K. Sahu, J. J. Valletta, C. D. Byrne, and A. J. Chipperfield, "A bayesian network for modelling blood glucose concentration and exercise in type 1 diabetes," *Statistical methods in medical research*, p. p. 0962280214520732, 2014.
- [17] W. P. Charette, *Control systems theory applied to metabolic homeostatic systems and the derivation and identification of mathematical models*. PhD thesis, California Institute of Technology, Pasadena, Calif, 1969.
- [18] R. O. Foster, "The dynamics of blood sugar regulation," Master's thesis, Massachusetts Institute of Technology, Cambridge, 1970.
- [19] W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter, and S. L. Pohl, "Evaluating clinical accuracy of systems for self-monitoring of blood glucose," *Diabetes Care*, vol. 10, p. 622–628, 1987.
- [20] J. Xie *et al.*, "Benchmarking machine learning algorithms on blood glucose prediction for type i diabetes in comparison with classical time-series models," *IEEE Transactions on Biomedical Engineering*, 2020.
- [21] S. M. Pappada *et al.*, "Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes," *Diabetes Technology & Therapeutics*, vol. 13, no. 2, 2011.
- [22] J. Martinsson *et al.*, "Blood glucose prediction with variance estimation using recurrent neural networks," *Journal of Healthcare Informatics Research*, vol. 4, pp. 1–18, 2020.

- [23] Y. Bengio *et al.*, “Machine learning for combinatorial optimization: a methodological tour d’horizon,” *Canada Excellence Research Chair in Data Science for Real-Time Decision-Making*, 2018.
- [24] S. Priyan, “Operations research in healthcare : A review,” *Juniper Online Journal of Public Health*, vol. 1, no. 3, 2017.
- [25] R. N. Bergman, Y. Ziya. Ider, C. R. Bowden, and C. Cobelli, “Quantitative estimation of insulin sensitivity,” *American Journal of Physiology-Endocrinology And Metabolism*, vol. 236, no. 6, p. p.E667, 1979.
- [26] M. Palumbo *et al.*, “Personalizing physical exercise in a computational model of fuel homeostasis,” *Computational Biology*, 2018.
- [27] J. KIM *et al.*, “Multi-scale computational model of fuel homeostasis during exercise: Effect of hormonal control,” *Annals of Biomedical Engineering*, vol. 35, no. 1, 2007.
- [28] F. K. Knop *et al.*, “Reduced incretin effect in type 2 diabetes: cause or consequence of the diabetic state?,” *Diabetes*, vol. 56, pp. 1951–1959, 2007.
- [29] G. Ahlborg and P. Felig, “Lactate and glucose exchange across the forearm, legs, and splanchnic bed during and after prolonged leg exercise.,” *J Clin Invest*, vol. 69, no. 1, pp. 45–54, 1982.
- [30] D. K. Rollins *et al.*, “Free-living inferential modeling of blood glucose level using only noninvasive inputs,” *Journal of Process Control*, vol. 20, pp. 95–107, 2010.
- [31] J. D. Elashoff, T. J. Reedy, and J. H. Meyer., “Analysis of gastric emptying data.,” *Gastroenterology*, vol. 83, p. 1306–12, 1982.
- [32] M. Berger and D. Rodbard, “Computer simulation of plasma insulin and glucose dynamics after subcutaneous insulin injection.,” *Diabetes Care*, vol. 12, pp. 725–36, 1989.
- [33] E. Børsheim and R. Bahr, “Effect of exercise intensity, duration and mode on post-exercise oxygen consumption,” *Norwegian University of Sport and Physical Education, Oslo, Norway*, 2003.
- [34] I. M. Schiavon, *Modeling The Effect Of Physical Activity On Postprandial Glucose Turnover*. PhD thesis, University of Padova, Department of Information Engineering, 2014.
- [35] C. C. Palerm and B. W. Bequette, “”hypoglycemia detection and prediction using continuous glucose monitoring—a study on hypoglycemic clamp data,” *Journal of Diabetes Science and Technology*, vol. 1, no. 5, 2007.

-
- [36] S. Oviedo, J. Vehí, , R. Calm, and J. Armengol, “A review of personalized blood glucose prediction strategies for t1dm patients,” *International journal for numerical methods in biomedical engineering*, vol. 33, no. 6, 2017.

Appendix A

Physiological Models of External Stimuli

A.1 Gastro intestinal track compartment

This compartment is responsible for producing glucose appearance rate in the circulation due to meal consumption. In order to model the process of glucose appearance from gastro intestinal track, the power exponential model by Elashoff et al. [31] is adopted.

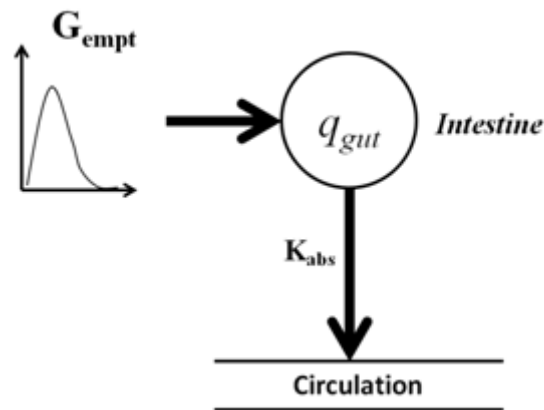


Figure A.1: Gastric emptying and appearance of glucose fluxes in circulation.

$$q'_{gut}(t) = -k_{abs} \cdot q_{gut}(t) + G_{empt}(t)$$

$$Ra(t) = f \cdot k_{abs} \cdot q_{gut}(t)$$

where,

q_{gut} is the amount of glucose in the gut,

k_{abs} is the rate constant of intestinal absorption ($k_{abs} = 1/40$)

f is the fraction of intestinal absorption appears in plasma.

These equations describe glucose absorption by the gut. It is assumed that the fraction of glucose in the duodenum increases following a power exponential function.

$$q_{duo}(t) = D \cdot \{1 - e^{-(kt)^\beta}\}$$

where,

k is the rate of emptying ($k = 1/40$)

β a shape factor ($\beta = 1, 2, 3, \dots$) means type of carbohydrate.

D is the amount of ingested carbohydrate in gram

Thus, the gastric emptying rate is

$$G_{emp}(t) = q'_{duo}(t) = D \cdot \beta \cdot k^\beta \cdot t^{\beta-1} \cdot e^{-(kt)^\beta}$$

A.2 Subcutaneous compartment

This compartment is responsible to produce continuous insulin appearance rate in case of subcutaneous insulin injection. In order to model the process of insulin appearance into blood circulation from the site of subcutaneous injection, the model proposed by Berger and Rodbard [32] is adopted. According to that model, subcutaneous injection is represented as below.

$$I_{abs}(t) = \frac{s \cdot t^s \cdot T_{50}^s \cdot D}{t \cdot [T_{50}^s + t^s]^2}$$

where,

t is the time elapsed from the injection,

D is the dose injected subcutaneously,

T_{50}^s is the time of absorption of 50% of the dose,

s is the preparation-specific parameter defining the insulin absorption pattern of the different types of insulin catered for in the model (regular, intermediate, lente and ultra lente).

A linear dependency of T_{50}^s on dose is defined as:

$$T_{50}^s = a * D + b$$

The values of a , b and s according to the *Berger and Rodbard* [32] are:

For regular/short acting insulin:

$$s=2.0; a=0.05; b=1.7;$$

For NPH/intermediate acting insulin:

$$s=2.0; a=0.18; b=4.9;$$

For lente/long acting insulin:

$$s=2.4; a=0.15; b=6.2;$$

For ultra-lente/very long acting insulin:

$$s=2.5; a=0; b=13;$$

A.3 Lungs compartment

Lungs compartment is responsible to produce respiratory oxygen consumption rate at basal and during physical activity. According to the model by A. Roy and R.S. Parker [13], capacity of an individual for aerobic work is indirectly measurable by the maximum oxygen consumption rate, VO_2^{max} . Percentage of maximum volume of oxygen consumption (PVO_2^{max}) can be used as the intensity level for quantification of exercise. PVO_2^{max} is increased rapidly at the onset of exercise and 5–6 minutes are taken for reaching to ultimate value and remains constant for the length of exercise. According to A. Roy and R.S. Parker, the PVO2max is represented as-

$$\frac{dPVO_2^{max}}{dt} = -0.8PVO_2^{max}(t) + 0.8 * u_3(t);$$

$$PVO_2^{max}(0) = 0;$$

Where, PVO_2^{max} is the exercise level as experienced by the individual, and $u_3(t)$ is the ultimate exercise intensity above the basal level. The parameter value of 0.8 (1/min) was selected to achieve a PVO_2^{max} settling time of approximately 5 minutes.

Appendix B

Real and Synthetic Dataset

Table B.1: OGTT dataset for proposed model validation.

Time(min)	Glucose(mmol/L)	Glucose(mg/dL)	Insulin(pmol/L)	Insulin(μ U/ml)
1	5.4	97.29828	38	5.472
5	5.2	93.69464		
10	6.1	109.91102	109	15.696
15	7.1	127.92922		
20	8.1	145.94742	230	33.12
30	9.1	163.96562	262	37.728
40	9.7	174.77654	302	43.488
50	9.1	163.96562	320	46.08
60	8.6	154.95652	296	42.624
70	7.6	136.93832	223	32.112
90	6.1	109.91102	136	19.584
120	4.9	88.28918	55	7.92
150	4.3	77.47826	25	3.6
180	4.3	77.47826	26	3.744
240	4.7	84.68554	20	2.88

Table B.2: Age, height, weight, maximum oxygen uptake and work load during prolonged exercise.

	Mean	SE	Range
Age, yr	26	0.7	20-31
Height, cm	182	1.4	169-187
Weight, kg	71	1.6	57-82
Maximum oxygen uptake, liters/min	3.8	0.13	2.6-4.8
Work Load, W	130	4.9	90-170

Table B.3: Mean values of arterial substrate and hormone concentrations during and after prolonged exercise.

Time (min)	Heart Rate (beats/min)	O₂ uptake (ml/min)	Glucose (mmol/liter)	Lactate (mmol/liter)	FFA (mmol/liter)	Glucagon (pg/ml)	Insulin μU/ml
1	52	283	4.39	0.57	0.43	77	14.5
40	141	2137	4.09	1.19	0.66	66	11.8
90	148	2269	3.86	1.31	0.83	111	9.2
120	148	2155	3.55	1.09	1.11	158	8
180	157	2305	2.78	1.56	1.92	257	6
210	150	2234	2.56	1.55	2.3	256.5	7.7
220	96	381	3.12	1.5	2.47	256	8.3
240	94	353	3.19	1.23	2.23	222	9
260	90	325	3.18	1	2	221	8
280	85	305	3.18	0.94	1.72	220	7.5

B.1 Algorithm for synthetic activity profile

In this algorithm, generation of synthetic activity profile as function of glucose variation is shown:

Step-1: Estimating glucose change rate:

$$R_t = (G_{t+1} - G_t) / G_t;$$

Step-2: Estimating activity intensity:

$$[PVO_2^{max}]_t = Basal_Rate * [EffectofGlucoseChange]_t;$$

$$[EffectofGlucoseChange]_t = [(a + b)/2 + (a-b)/2 * \tanh(c * (R_t - x_0 + d))];$$

Step-3: Estimating parameter through nonlinear curve fitting:

$$a = 6; b = -5; x_0 = -0.01; c = 28.4967463; d = 0.05966283;$$

Table B.4: Assumed dataset for establishing relation between exercise and glucose variation.

Glucose Change Rate	Effect
-0.01	1
-0.03	2
-0.05	4
-0.07	6
-0.08	7
-0.1	9.5
-0.12	11
-0.14	11
-0.16	11
-0.18	11
-0.2	11

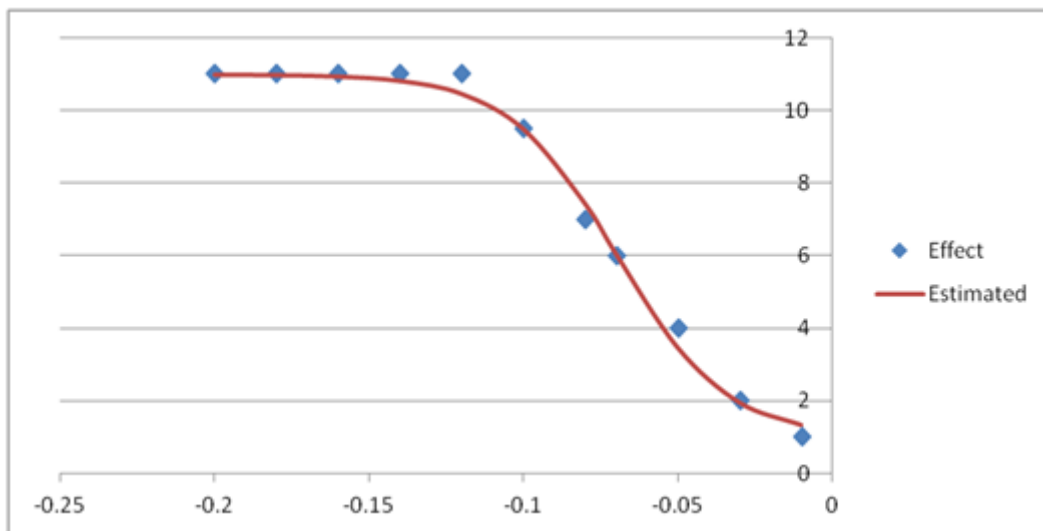


Figure B.1: Hyperbolic tangential relation between exercise and glucose excursion.

Appendix C

Simulink Implementation of IPM

The proposed integrated physiological model (IPM) of this thesis is implemented on MATLAB/Simulink Environment and all optimization is done using simulink design optimization tool. In this part of the document, implementation views of different compartment are shown.

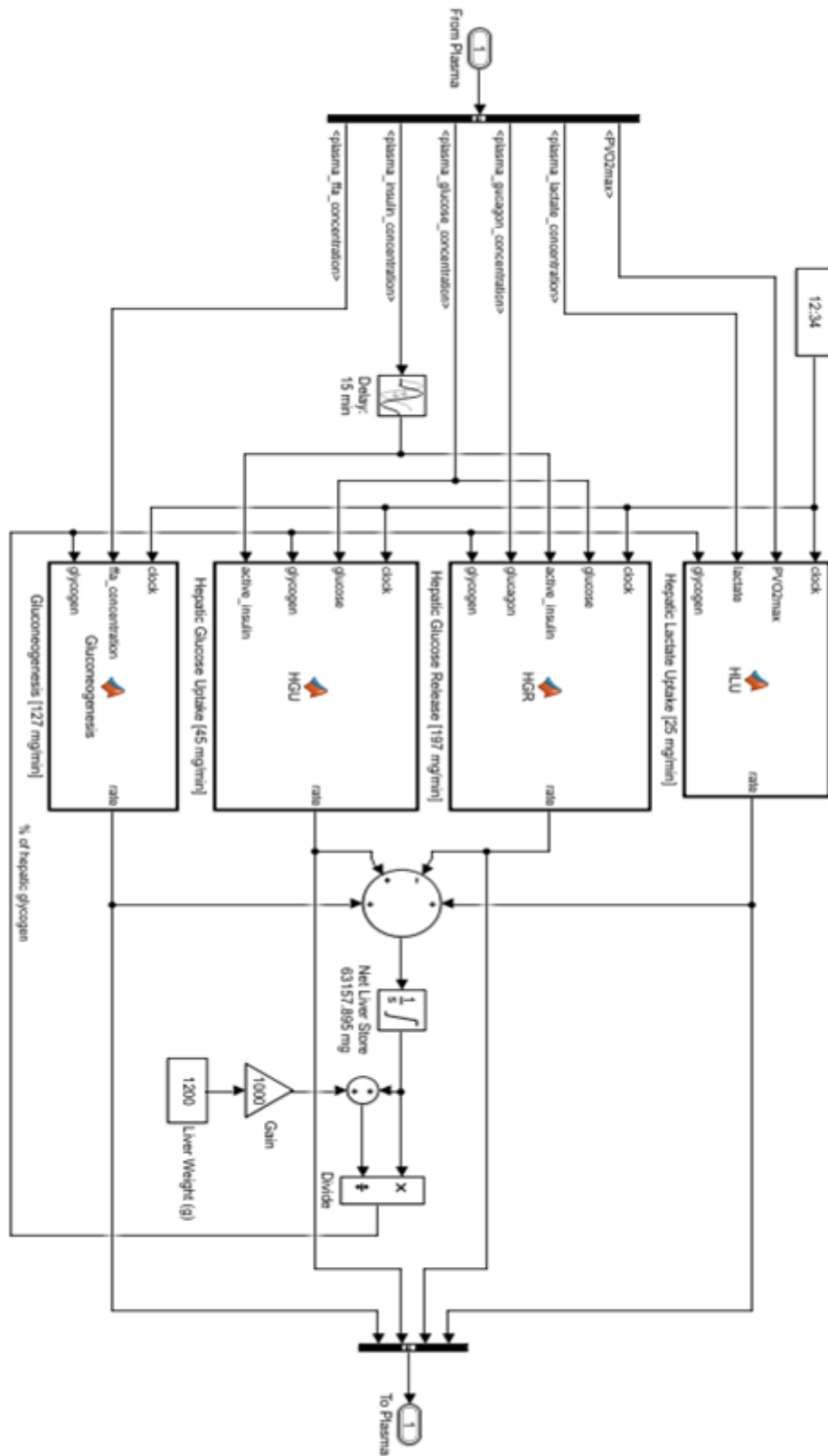


Figure C.1: Simulink implementation of liver compartment.

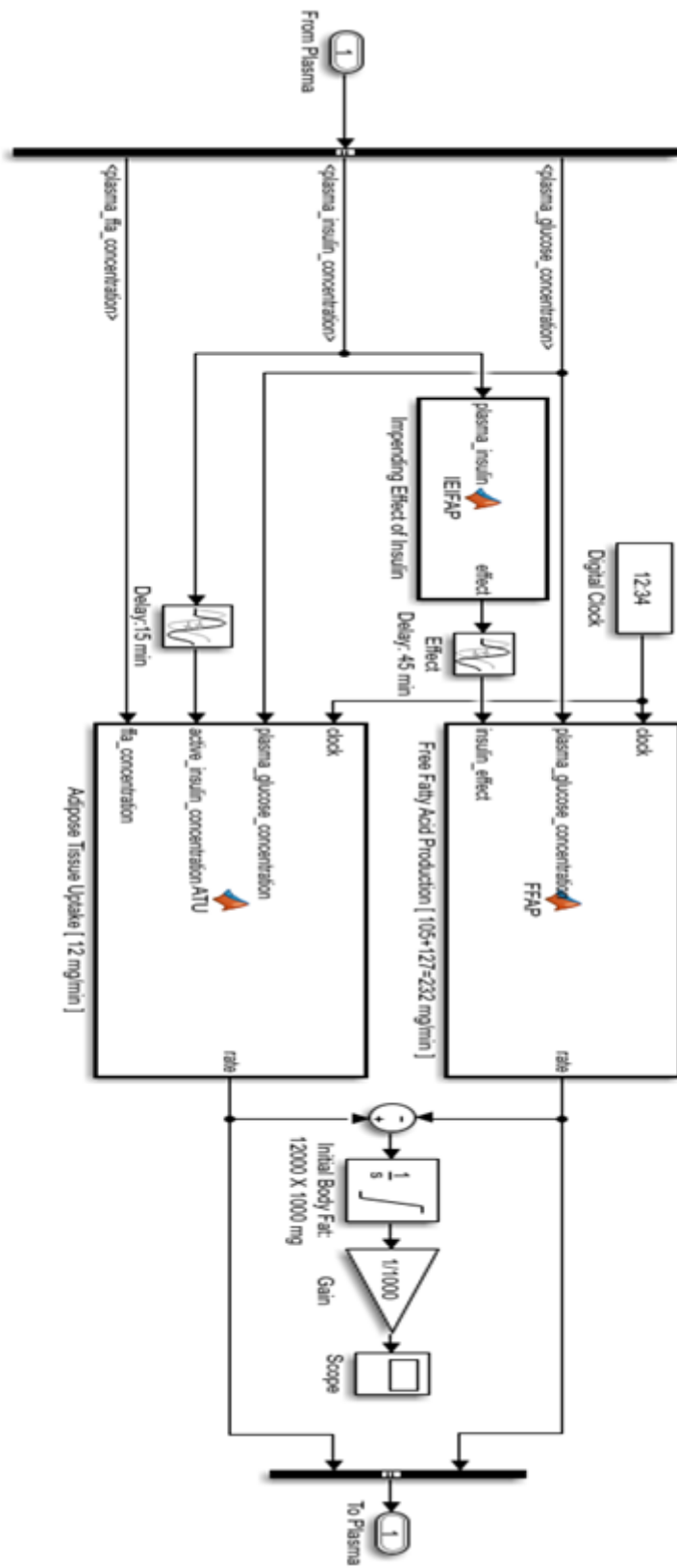
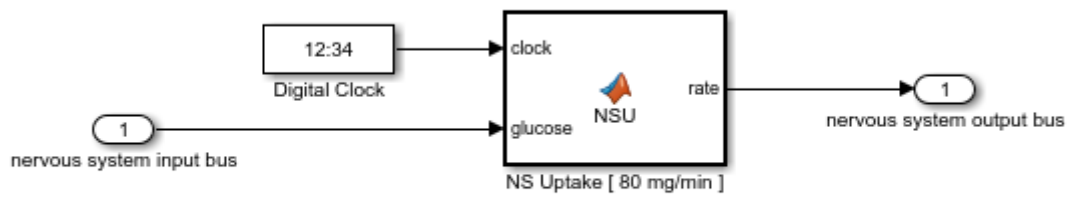


Figure C.3: Simulink implementation of adipose compartment.



the glucose taken up by the nervous system is completely burned

Figure C.4: Simulink implementation of nervous system compartment.

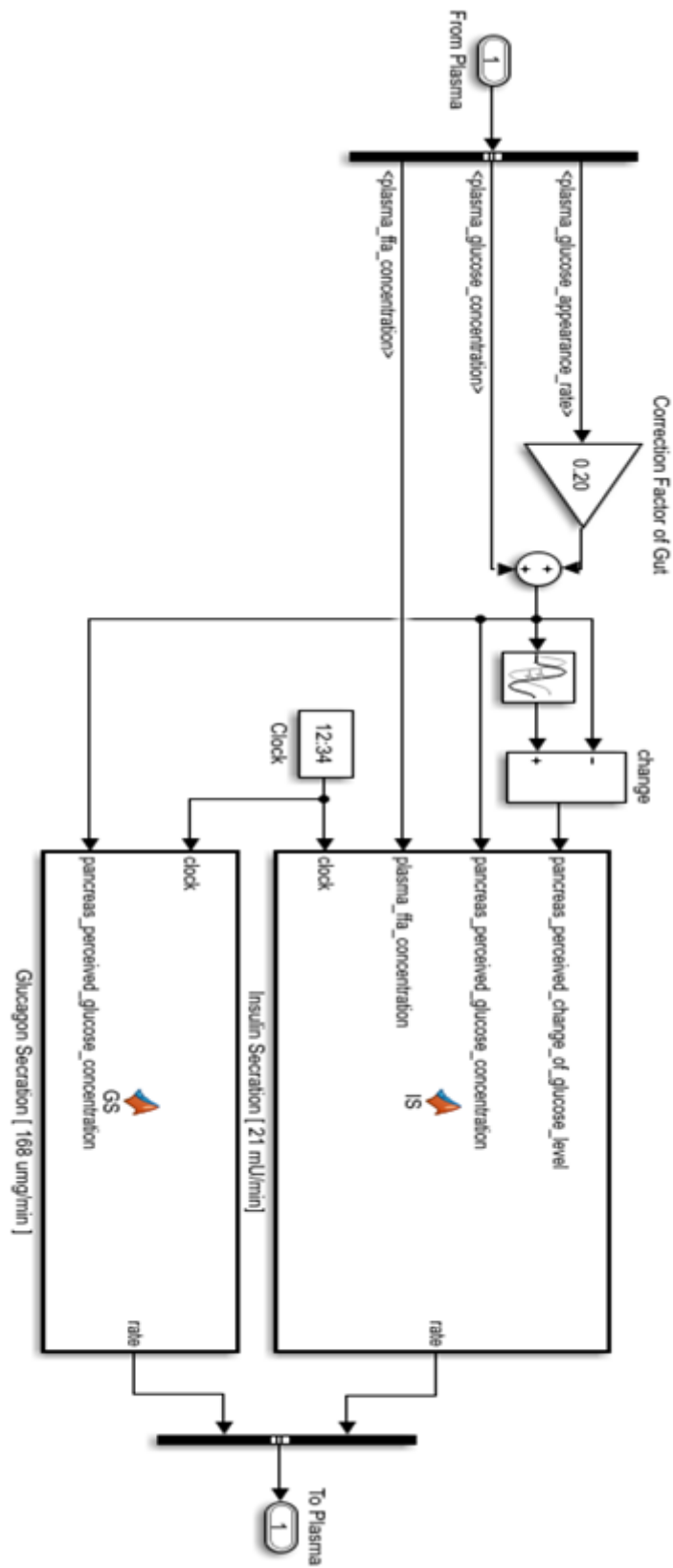


Figure C.5: Simulink implementation of pancreas compartment.

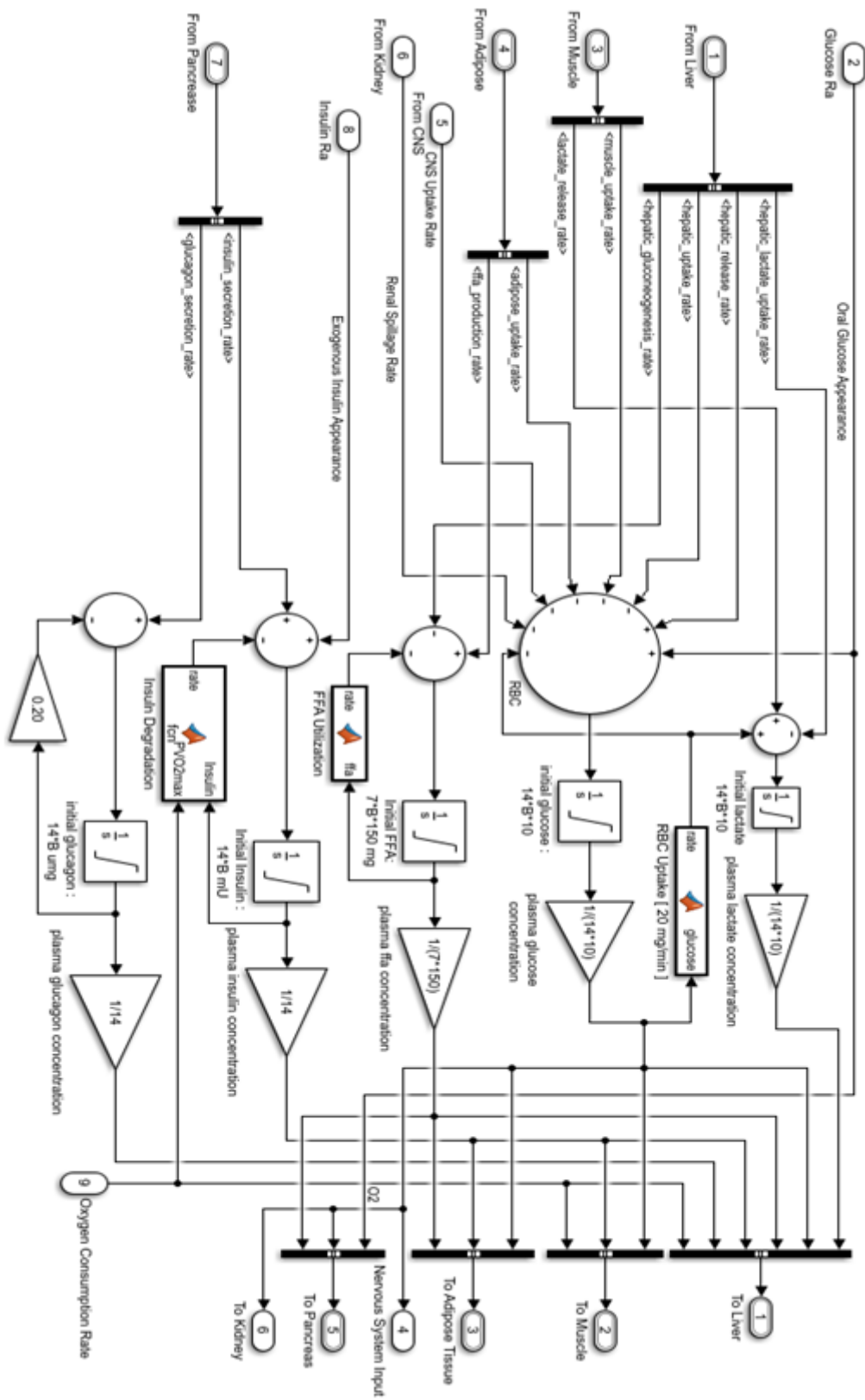


Figure C.6: Simulink implementation of plasma compartment.

Appendix D

Thesis Reproduction

The materials including all data, model files and scripts to reproduce the computational experiments of the thesis are released under an open-source license in Mendeley Data. We have used Matlab (R2017a) for writing scripts, simulink for implementing models, and Neural Network Toolbox (10.0) for building NN models. The url of the repository is DOI: [10.17632/gb5bd386g4.2](https://doi.org/10.17632/gb5bd386g4.2)

Generated using Postgraduate Thesis L^AT_EX Template, Version 1.03. Department of
Computer Science and Engineering, Bangladesh University of Engineering and
Technology, Dhaka, Bangladesh.

This thesis was generated on Sunday 28th March, 2021 at 8:53am.