HAW HAMBURG

MASTERTHESIS
Hannes Voß

Controlling Insulin Delivery in Type 1 Diabetes: A Deep Reinforcement Learning Approach

FAKULTÄT TECHNIK UND INFORMATIK
Department Informatik

Faculty of Computer Science and Engineering
Department Computer Science

HOCHSCHULE FÜR ANGEWANDTE WISSENSCHAFTEN HAMBURG
Hamburg University of Applied Sciences

Hannes Voß

Controlling Insulin Delivery in Type 1 Diabetes: A Deep Reinforcement Learning Approach

Masterarbeit eingereicht im Rahmen der Masterprüfung
im Studiengang *Master of Science Informatik*
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Olaf Zukunft
Zweitgutachter: Prof. Dr. Zhen Ru Dai

Eingereicht am: 02. Februar 2022

Hannes Voß

Thema der Arbeit

Steuerung der Insulinabgabe bei Typ-1-Diabetes: Ein Deep Reinforcement Learning-Ansatz

Stichworte

Diabetes, Insulinabgabe, Deep Reinforcement Learning, Artificial Pancreas, Machine Learning, Closed Loop

Kurzzusammenfassung

Viele Forschende verfolgen das Ziel, Diabetikern eine interventionsfreie Diabetes-Therapie zu ermöglichen. In den letzten Jahren hat die Entwicklung einer technischen künstlichen Bauchspeicheldrüse immer mehr an Bedeutung gewonnen. Fortschritte im Verständnis von Diabetes und eine vollständige Überwachung des Blutzuckerspiegels tragen wesentlich zum Fortschritt bei. Um die Entwicklung einer künstlichen Bauchspeicheldrüse zu unterstützen, ist es notwendig, sich mit der Entwicklung eines Algorithmus zu beschäftigen, der die Insulinabgabe in Abhängigkeit von den Stoffwechselfaktoren eines Patienten steuert. In dieser Arbeit wird die Entwicklung und in silico Evaluation eines Algorithmus zur Insulinabgabe auf Basis von Deep Q-Learning vorgestellt. Zunächst wird ein PID Controller vorgestellt und evaluiert. Darauf folgend wird untersucht, wie sich das Skalieren des Eingabevektors auf die Time-In-Range (TIR) eines simulierten Patienten auswirkt. Zudem wird untersucht, ob die Verwendung von Insulin-On-Board (IOB) als Parameter für die Insulinabgabe geeignet ist. Es wird auch evaluiert, wie unterschiedliche Konfigurationen des Action-Spaces und der Sequenzlänge die Time-In-Range (TIR) der simulierten Patienten verbessern. Das Skalieren des Eingabevektors führt zu einer Veränderung der Time-In-Range (TIR) von $52.24\% \pm 15.81$ zu $50.62\% \pm 12.9$. Unter Einbezug von Insulin-On-Board (IOB) zeigen die Ergebnisse eine deutliche Verbesserung der Time-In-Range (TIR) der simulierten Patienten von $50.62\% \pm 12.9$ zu $57.71\% \pm 12.12$. Die Evaluation eines größeren Action-Spaces zeigt eine Reduktion der Standardabweichung der Time-In-Range (TIR) von $57.71\% \pm 12.12$ zu $57.57\% \pm 7.18$. Darüber hinaus wird ein Ausblick gegeben, welche weiteren auf dieser Arbeit aufbauenden Forschungsaktivitäten zur kontinuierlichen Entwicklung einer künstlichen Bauchspeicheldrüse geplant sind.

Hannes Voß

Title of Thesis

Controlling Insulin Delivery in Type 1 Diabetes: A Deep Reinforcement Learning Approach

Keywords

Diabetes, Insulin delivery, Deep Reinforcement Learning, Artificial Pancreas, Machine Learning, Closed Loop

Abstract

Many researchers are pursuing the goal of enabling intervention-free diabetes therapy for diabetics. In recent years the development of a technical artificial pancreas has become more and more important. Advances in understanding diabetes and a complete monitoring of the blood sugar level contribute significantly to progress. To support the development of an artificial pancreas it is necessary to deal with the development of an algorithm that controls the insulin delivery depending on the metabolic factors of a patient. In this thesis the development and in silico evaluation of an insulin delivery algorithm based on Deep Q-Learning is presented. First, a PID controller is presented and evaluated. Subsequently, it is investigated how scaling the input vector affects the Time-In-Range (TIR) of a simulated patient. In addition, the suitability of using Insulin-On-Board (IOB) as a parameter for insulin delivery is investigated. It will also be evaluated how different configurations of action space and sequence length improve the Time-In-Range (TIR) of simulated patients. Scaling the input vector leads to a change in Time-In-Range (TIR) from $52.24\% \pm 15.81$ to $50.62\% \pm 12.9$. Including Insulin-On-Board (IOB), results show a significant improvement in Time-In-Range (TIR) of simulated patients from $50.62\% \pm 12.9$ to $57.71\% \pm 12.12$. Evaluation of a larger action space shows a reduction of the standard deviation in Time-In-Range (TIR) from $57.71\% \pm 12.12$ to $57.57\% \pm 7.18$. Furthermore, an outlook is given on which further research activities based on this work are planned for the continuous development of an artificial pancreas.

Contents

List of Figures

List of Tables

1 Introduction

There are currently more than 422 million people suffering from diabetes worldwide [101]. People suffering from diabetes have to plan and carry out the therapy of their disease attentively and meticulously throughout their lives, which for many represents an immense burden and reduction in the general quality of life [20]. It has been known since 1983 that long-term therapy-induced stress can lead to what is known as diabetes burnout [42, 78]. This constant stress on a person with diabetes (sometimes called diabetes distress) also increases the risk of developing depression [83]. A notable prevalence of depressive symptoms (up to 30%) and major depression (11-13%) has been observed in this context [8, 2]. Suffering from treatment-induced depression may, in turn, lead to inconsistent compliance with diabetes therapy, creating a mutually dependent cycle [32].

Fortunately, in the field of diabetes management, a number of technical innovations have come on the scene in recent years. Although these innovations make it easier to regulate the disease, many manual patient actions are still necessary to regulate the disease. To enable diabetes therapy to be as intervention-free as possible more processes need to be automated. The idea of developing an artificial pancreas that automates the manual patient actions has existed since the 1960s. In recent years the idea of a software-supported artificial pancreas has become increasingly popular [25, 75, 11].

1.1 Scope

The goal of this thesis is to answer whether Deep Q-Learning is suitable to solve the insulin delivery control problem. It will be investigated whether the differences in the metabolism of the patients are recognized, learned and answered adequately. Also, the question whether regular eating behavior with high-carb meals and deviations from it will be handled with an adequate control strategy will be answered. It should also be investigated if an approach to be developed can prevent dangerous glycaemic events.

Out of scope of this thesis is the development of an approach which is considered as a medical device. The results of this work are therefore not intended to be used in the context of or as a medical device and do not claim to fulfill medical regulations.

From these preliminary thoughts, the following specific research questions can be derived:

RQ 1 How does a Deep Q-Learning approach needs to be designed to meet the requirements of good diabetes therapy solely through basal rate adjustment despite the different metabolic factors of individual patients?

RQ 2 To what extent can a good diabetes therapy be ensured despite deviations from regular eating patterns?

RQ 3 How can the approach act safely within predefined boundaries and prevent severe hypoglycemia?

RQ 4 What is the impact of using Insulin-On-Board (IOB) as a factor of the algorithm to be developed?

RQ 5 How effectively can the approach being developed handle high-carb meals?

In order to answer these questions adequately, the current state of research in the research area of insulin delivery algorithms will first be investigated and the most relevant approaches will be compared according to defined criteria. Subsequently, in silico experiments will be designed and conducted with a Deep Q-Learning algorithm to be developed using the UVa/Padova T1D simulator. The results will then be evaluated and discussed to answer the presented research questions. In the end result, this thesis presents an insulin delivery algorithm based on Deep Reinforcement Learning, which will contribute to the advancement of technologies for the development of an artificial pancreas.

1.2 Outline

This thesis is organized as follows. Chapter 2 begins by explaining the basic principles and terminology necessary for a fundamental understanding of this thesis. Chapter 3 deals with the current state of research. Chapter 4 discusses the eligible data resources to conduct the experiments. Chapter 5 explains the experiment setup for the experiments to be conducted. Chapter 6 presents the experiments in detail. In Chapter 7, the results

obtained get evaluated and placed in the context of current research in comparison. Chapter 8 briefly summarizes the results of the work and provides an outlook on the further research steps planned to follow on this work.

2 Background

In order to fully understand the content of this thesis, a few basic terms will be clarified in this chapter. First, the disease diabetes mellitus is described in its basic characteristics. Here, the consequences of inadequate therapy and the most basic aids in modern diabetes therapy are presented. Then the essential components of an Artificial Pancreas System, their different characteristics and the benefits of such a system for the patient are described. Next, the concept of reinforcement learning is explained and the potential benefits for diabetes therapy are presented. In the last section, a few needed domain specific terms will be clarified.

2.1 Diabetes mellitus

Diabetes is a chronic metabolic disease of the pancreas which, if not treated, leads to an acute insulin deficiency in the body. It is divided into several types, each of which can have different triggers. However, all types have in common that the glucose content of the blood must be kept within a target range of 70 - 180 mg/dL. In type 1 diabetes the body's immune system destroys the beta cells responsible for the body's insulin production [46]. As soon as they are no longer present, it is no longer possible to produce enough insulin to utilize the carbohydrates absorbed. This insulin deficiency must be compensated by a constant subcutaneous supply of insulin. However, a dosage of insulin that is too high leads to hypoglycemia (concentration below 70 mg/dL), while a dosage that is too low can lead to hyperglycemia (concentration above 180 mg/dL). The target range between 70 mg/dL and 180 mg/dL is also called *euglycemic range*. However, the term is not defined in a standardized way and may be interpreted differently depending on the study [31]. A longterm underdose means that the glucose in the blood can no longer be released into the body's cells and the body must access the body's fat reserves as a source of energy. In the long term, this leads to poisoning of the body with ketone bodies (ketoacidosis), which can lead to coma and, in the worst case, death. A longterm wrong dosage with long

periods of glucose concentration above 180 mg/dL can additionally trigger a number of secondary diseases. These include diabetic retinopathy, kidney damage and circulatory disorders [26].

Since the introduction of *Continuous Glucose Measurement (CGM)* systems, glucose concentration can be continuously measured from the interstitial fluid at intervals of typically 3-5 minutes and transmitted to a receiving device (such as a smartphone). This development not only leads to a significant improvement in the therapy parameters of a diabetic patient [38] but also enables a much deeper understanding of the disease, as data is now collected at small regular intervals instead of just snapshots from blood glucose measurements, as was previously common practice. This data can be used for research purposes with the patient's consent. However, the glucose concentration in the interstitial fluid differs from the blood sugar concentration. It must be assumed that the values measured with a CGM have an estimated time difference of about 4 to 26 minutes to the data measured from the blood which should be considered when developing an insulin delivery algorithm [72].

Another development heralding a new era in patient diabetes management was the first official insulin pump Promedos from Siemens launched in 1981 [5]. By delivering small amounts of insulin per hour (the so-called basalrate) it is possible to keep the glucose concentration in the blood constant in the patient's fasting state and to reduce the risk of body poisoning caused by insulin deficiency (ketoacidosis) [76]. However, it is still necessary to manually deliver a higher amount of insulin when taking in carbohydrates (called bolus insulin) in order to compensate for the increase in blood sugar caused by them.

2.2 Artificial Pancreas

A so-called *Artificial Pancreas* (sometimes referred to as *loop system*) consists of several components with different functions. Typically, these components are an insulin pump, a continuous glucose monitoring system, the patient themselves, and a feedback control system that calculates and delivers the insulin to be delivered based on the patient metabolic parameters measured by the sensor component (see Fig. 2.1) [86].

A distinction is made between a *Closed Loop* and a *Hybrid Loop* system. In a closed loop system, blood glucose is completely regulated by the delivery of insulin (at higher

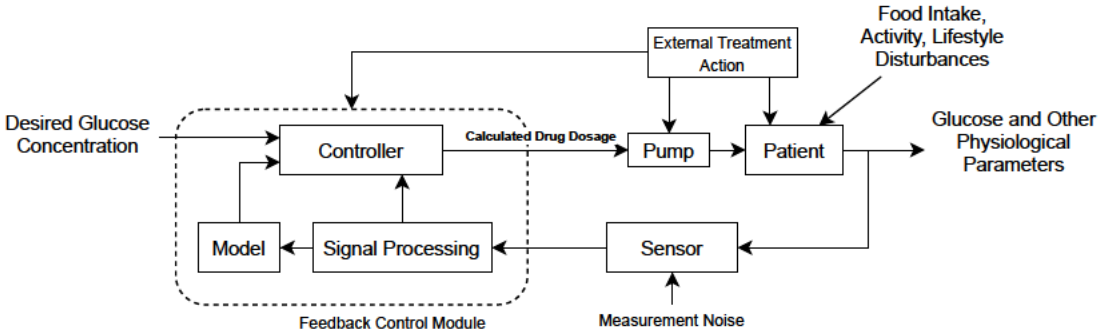

Figure 2.1: Components of an Artificial Pancreas [86].

glucose concentrations) and/or glucagon (at lower glucose concentrations). No further human intervention is necessary. In a hybrid closed loop system, glucose concentration is automatically regulated by insulin delivery as in a closed loop system, but the patient still has to support the system with manual input. Meals, for example, must be announced and manually controlled so that the loop can react appropriately. At the moment, hybrid closed loop systems are mostly only available as self-made systems with non-mature algorithms [34]. To date, only Medtronic has developed an approved hybrid closed-loop system, the MiniMed^{TM} 670G/780G [54]. The current version of this commercial model shows significant improvements in a patient's blood glucose levels [27].

In general many study results show that Artificial Pancreas systems significantly improve diabetes therapy [19, 10, 99, 45, 28]. However, the systems tested are hybrid loop systems and still require a lot of manual intervention to provide optimal therapy [19].

Another distinction of Artificial Pancreas algorithms is the choice between mono- and bihormonal control algorithms [86]. Research is currently uncertain whether a monohormonal or a bihormonal system is suitable for safe everyday use. Strong research is being conducted on both approaches [39, 6, 99]. The use of bihormonal systems includes the calculation and delivery of glucagon in addition to the delivery of insulin. The administration of glucagon generates glucose in the body and consequently causes the glucose concentration in the blood to rise. In this way, severe hypoglycemia in particular can be avoided or compensated for. Comparison of the two systems shows that a dual-hormone system reduces the risk of daytime hypoglycemia, especially during exercise, whereas single-hormone systems appear to handle nighttime hypoglycemia better [39]. However, the glucagon currently available on the market is not yet very mature and still has major

problems with its formulation and can have a toxic effect in a patient's body. Therefore, glucagon needs to be further developed for safe use in bihormonal pumps [6].

From these preliminary considerations, it has been decided that this thesis will focus on the development and evaluation of a monohormonal insulin delivery algorithm.

The research area of Artificial Pancreas is divided into different sub-disciplines, of which five sub-disciplines relevant to this thesis are presented below:

1. Low Glucose Suspension,
2. Logic-based Control,
3. Proportional Integral Derivative Control,
4. Model-Predictive Control,
5. Machine-Learning based approaches [86].

2.2.1 Low Glucose Suspension

Probably the simplest Artificial Pancreas algorithms imaginable are those of the *Low Glucose Suspension (LGS)* category. When leaving the euglycemic range, the insulin supply is either stopped (if the glucose concentration is too low) and restarted when the target value is reached or the desired target range is exceeded [86].

In the *Predictive Low Glucose Suspension (PLGS)* algorithms, the future glucose concentration is predicted to initiate the shutdown of insulin delivery before hypoglycemia occurs [35].

2.2.2 Logic-based Control

A further class of Artificial Pancreas algorithms are the so-called *Logic-based Control* algorithms. The most well-known controller of this class is the fuzzy controller. A fuzzy controller is a controller that works with fuzzy process information [86].

The fundamental steps in a fuzzy controller include [84, 73]:

1. Input fuzzification,

2. Fuzzy inference,

3. Output defuzzification.

2.2.3 Proportional Integral Derivative Control

A very common controller type in the context of Artificial Pancreas is the *Proportional Integral Derivative (PID)* controller. It is a controller type from Control Theory that works by computing the past-present-future error of an action in a control feedback problem [86]. In the context of Artificial Pancreas, insulin delivery is adjusted by assessing the deviation from the target glucose level (the proportional component), the area under the curve between the measured and target glucose concentrations (the integral component), and the rate of change in the measured glucose concentration (the derivative component) [95].

PID controllers are by far the most extensively researched control type in the context of Artificial Pancreas [88].

2.2.4 Model Predictive Control

Another approach in the field of control engineering and the research area of Artificial Pancreas is the so-called *Model-Predictive Control (MPC)*. Rather than an algorithm, it is more of a general approach to solving control problems [13]. It is based on a discrete-time dynamic model of the process to be controlled, which is used to predict future behavior of a process as a function of input signals [87].

Model predictive control is also receiving much attention in the context of insulin delivery algorithm development [31, 13].

2.3 Reinforcement Learning

A further class of control algorithms that is fundamentally important for this thesis is *Reinforcement Learning*. Reinforcement Learning is a sub-discipline of Machine Learning, which is a scientific discipline that allows computer systems to develop intelligent behavior from empirically collected data. Its relevance to the medical field has grown rapidly in

recent years and is attracting more and more attention in healthcare research [44]. Unlike the disciplines *Supervised Learning* and *Unsupervised Learning*, the computer system learns based on its own experience in a dynamic system. No detailed description of the environment or labeled data is required. Reinforcement learning is especially suitable for systems with inherent time delays. This makes reinforcement learning particularly interesting for the use in the context of insulin delivery, since, as already described, delays such as CGM delay and insulin pharmacodynamics have an impact on the success of therapy [18]. Other advantages of reinforcement learning in this context are that other data streams (such as other acquired sensor data from the patient) can be easily added. Minimal assumptions are made about the structure of the underlying process, which allows the algorithm to adapt to different metabolisms from different patients or to changes in a patient's metabolism over time. It also allows learning of regularities in meal intake, which is important when developing an insulin delivery algorithm [37].

2.3.1 Markov Decision Process

To understand the fundamentals of reinforcement learning it is crucial to understand, that it makes use of the formal definition of a *Markov Decision Process (MDP)*. A Markov Decision Process is a mathematical problem in which an agent's benefit depends on a sequence of decisions. A problem is considered as Markovian if it fulfills the so-called *Markov Property*. In a simplified way, this means that the probability of a state transition from state s to s' depends only on its prior state s and not on any other previously seen states [92].

Reinforcement learning problems are described as MDPs and operate in an *environment* that is fully or only partially observable, depending on the problem. The *states* s of the environment in general contain information about the problems environment.

In particular, the control problem of correct insulin delivery can be described as an infinite Markov Decision Process in the form of a tuple of (S, A, T, R, γ). Since the metabolism of a patient cannot be fully observed in all its parameters, the problem is considered as partially observable. This implies that recurrent units should be used in the neural network to be created [92].

One run of an algorithm is referred to as an *episode*, which, depending on the problem to be solved, ends after a specified length (timesteps) or continues to run until a termination criterion is met (usually the failure of the algorithm). Within an episode, the agent

receives the state of the environment and executes an *action* a based on the interpretation of the results. After executing the action, the agent receives a *reward* r, which can be interpreted as an evaluation of the action. The goal of these algorithms is to maximize cumulative rewards over time. It is important to note that the reward can be both positive and negative to reward or penalize decisions [92].

To revisit the definition of MDP the set of states (S) is composed of relevant measured values of the patient. Different experiments will be conducted with different amounts in the state S_t, but what they all have in common is the glucose concentration G measured at timestep t.

The set of actions (A) corresponds to the executable insulin delivery steps, while the state transition function (T) defines the probability of transitioning from state s to state s' when action a is executed while reward r is obtained.

To let the algorithm learn, the environment is first explored randomly. The randomness (often called epsilon ϵ) is started at a start value, which after a certain specified function runs against a fixed end value (close to 0). This behavior is called *epsilon greedy* and causes the algorithm to try many possible actions at the beginning and learn from the results [92]. The strategy learned over time is also called *policy* (π) [92].

In order to be able to evaluate whether a performed action was good or bad to achieve the defined goal, the value must be defined. However, the action alone indicates nothing about the value of the action - it also always depends on the state in which the environment is present. This evaluation of the state-action pair is also called *Q-Value* [92].

2.3.2 Q-Learning

In Q-Learning, the mapping of a state-action pair to a Q-Value takes place in the so-called *action-value-function*, also called Q-function (see Eq. 2.1).

$$Q^{\pi}(s,a) = \mathbb{E}\left[\sum_{t'=t}^{\infty} \gamma^{t'-t} r_{t'} \mid s_t = s, a_t = a, \pi\right]. \tag{2.1}$$

In Eq. 2.1, the Q-Value for a state-action pair is defined over the expected value of the accumulated and discounted rewards when following policy π. The discount factor

$\gamma \in [0, 1]$ ensures that when the choice is less than 1, rewards far in the future are weighted less than those immediately ahead. It is set relatively high in this context, because the dynamics of insulin persists in the body for a relatively long period of time.

In order to be able to maximize the rewards signal over time, i.e. to calculate the optimal Q-Value, the Bellman optimality equation is used (see Eq. 2.2).

$$Q^*(s, a) = \mathbb{E}_{s'} \left[R(s, a) + \gamma \max_{a'} Q^*(s', a') \right]. \tag{2.2}$$

The optimal Q-Value for the current state s is obtained here by selecting the action that maximizes the expected reward with the optimal $Q^*(s', a')$ in the next state s'. The results get stored in a table called the *Q-table* [92].

In Deep Q-learning a neural network is used instead of the Q-table to approximate the Q-Values.

2.4 Domain background

In order to understand the domain specific terms which are used in this thesis, some domain background information will be explained in this section. For this purpose, the concept of $\mathrm{HbA_{1C}}$ and the concept of Time-In-Range are first explained and the relevance for quality measurement in modern diabetes therapy is clarified. Furthermore, the concepts of the Blood Glucose Risk Index, Control-Variability Grid Analysis, and the pharmacodynamics of insulin are briefly described. The last section then presents the most important factors of influence on a patient's metabolic control and demonstrates the importance of this information.

2.4.1 $\mathrm{HbA_{1C}}$

To answer the proposed research questions, it is necessary to define what constitutes a good diabetes therapy. For this purpose, the so-called *HbA1C* is primarily used. The term is composed of hemoglobin (the pigment of red blood cells) and $\mathrm{A_{1C}}$, a glucose-binding protein chain. The percentage of bound glucose is usually determined by blood sampling to assess the long-term effectiveness of diabetes therapy. Since red blood cells are regularly produced in the spinal cord of the body, the $\mathrm{HbA_{1C}}$ roughly reflects the

metabolic state of a period of two to three months. The HbA_{1C} is usually given as a percentage, which in a healthy person is usually less than 6%. In untreated diabetes mellitus, the HbA_{1C} may be above 15%. In treated diabetes mellitus, the target HbA_{1C} which corresponds to a good diabetes therapy is 6-8% [57].

2.4.2 Time-In-Range

Another important metric in the evaluation of diabetes therapy is the so-called *Time-In-Range (TIR)*. This metric is used to evaluate both short-term and long-term metabolic adjustment. It is usually expressed as a percentage and indicates the time period in which the measured values are within the euglycemic range of 70-180 mg/dL [1].

The *Time-Above-Range (TAR)* describes the percentage of time the glucose concentration has been above the euglycemic range, while the *Time-Below-Range (TBR)* describes the percentage of time the glucose concentration has been below the euglycemic range [1].

It is possible to derive a rough HbA_{1C} from the TIR. If a TIR of at least 50% is achieved, the HbA_{1C} is around 7%, which corresponds to the goal of a successful diabetes therapy [21].

2.4.3 Blood Glucose Risk Index

The so-called *Blood Glucose Risk Index (BGRI)* is an index that describes the risk of a measured glucose concentration. It was first described by Kovatchev et al. [50]. The calculated risk index returns negative values at very low glucose concentrations while it is correspondingly high at very high glucose concentrations. It is utilized to define the two metrics described below.

The *Low Blood Glucose Index (LBGI)* and the *High Blood Glucose Index (HBGI)* are glycaemic control markers which describe the average corresponding risk-function return values over all glucose measurements x of size i. The function $rl(x_i)$ returns 0 for a measured glucose value if the BGRI function is positive for this value or the calculated negative value. For the function $rh(x_i)$ the same applies, except that this time only the positive risk index values are returned. They are defined as Eq. 2.3 and Eq. 2.3 show.

$$LBGI = \frac{1}{n}\sum_{i=1}^{n} rl(x_i) \tag{2.3}$$

$$HBGI = \frac{1}{n}\sum_{i=1}^{n} rh(x_i) \tag{2.4}$$

2.4.4 Control-Variability Grid Analysis

The *Control-Variability Grid Analysis (CVGA)* is a tool intended for quality measurement of a closed-loop algorithm on a group of patients. It plots extreme variations in glucose concentration over a given time period on a 3×3 risk matrix. Each patient is located on the matrix using a data point [62]. It will be used during the evaluation of the proposed approach.

2.4.5 Pharmacodynamics of insulin

To better understand why the development of insulin delivery algorithms is such a challenge, it is necessary to first look at the pharmacodynamics of insulin. Rapid-acting insulins, such as those used in insulin pumps, usually reach their maximum effect after 90 minutes and often act in the body for between six and eight hours (see Fig. 2.2) [12].

This is important to be aware of, because when calculating the amount of insulin to be delivered, the amount of insulin that may still be effective in the body (also called *insulin on board*) must be taken into account [12].

2.4.6 Factors of influence

Glucose levels can be affected by many different factors that are difficult to predict. The most important and influential are the so-called *MESS* influencing factors. MESS stands for Meals, Exercise, Stress and Sleep [24]. The ingested food has a different effect in every body. The insulin sensitivity of a patient, for example, changes depending on the patient's weight or age. At a higher weight, more insulin is often needed for the same amount of ingested carbohydrates than at a lower weight [24]. Meals remain one of the biggest

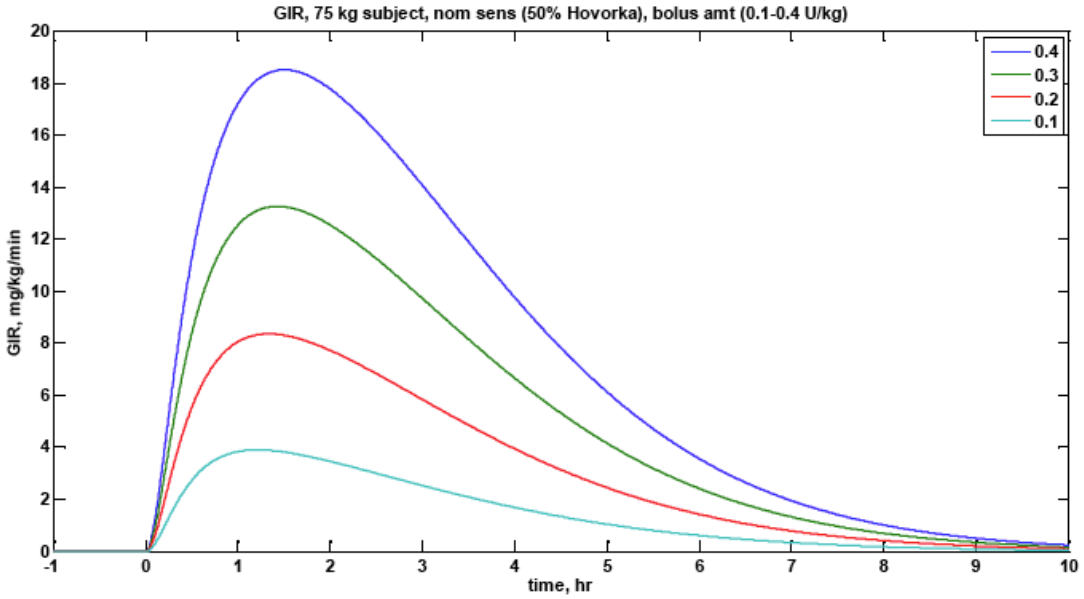

Figure 2.2: Pharmacodynamic profiles for four different bolus magnitudes (0.1-0.4 U/kg) of rapid-acting insulin [12].

challenges in the development of closed-loop algorithms [81]. But sports, stress and sleep are also difficult to assess and have different effects on glucose concentration depending on the patient [12]. One example is the so-called *dawn phenomenon* in which many patients experience a rise in glucose levels in their sleep during the morning hours around 04:00 am. This reduction in insulin sensitivity is differently pronounced in patients [12].

It is therefore important to understand that diabetes is a very individual disease that also requires an individual therapy and therefore an individual treatment in the algorithms to be developed. These characteristics lead to the assumption that the problem of multifactorial insulin delivery can be considered as a reasonable candidate for reinforcement learning.

3 Related Work

In order to understand the research area of insulin delivery algorithms, this chapter will present relevant related approaches. In addition, in cases where the authors have evaluated their approaches, their findings are also briefly presented here. The approaches that are most comparable with each other are summarized in tabular form to allow subsequent comparison of the presented approach in the research area.

The use of PID controllers is popular among various research groups in the Artificial Pancreas research area [88, 99]. Many of these algorithms have already been evaluated in clinical trials in patients [51, 85, 16, 23, 17, 58, 60, 59, 71].

A very well known PID controller is the controller presented by Steil et al. [89]. In this approach, the authors evaluate the algorithm over about 30 hours in a clinical setting with ten patients. They describe their results with a Time-In-Range (TIR) of 75%. The approach does not use meal boluses and thus belongs to the closed-loop control category.

Another closed-loop approach by Rossetti et al. [81] presents an algorithm based on a PID controller. Their approach limits the Insulin-On-Board (IOB) amount with a sliding mode reference conditioning (SMRC) procedure, which turns off the insulin delivery when the IOB is too high [79]. The authors evaluate their algorithm in a clinical setting on 20 patients and achieve a Time-In-Range (TIR) of 78.8%. The Time-Below-Range (TBR) is presented as 5.2%.

An approach developed by Bamgbose et al. [7] uses a system divided into two subsystems that uses neural network-based glucose prediction to calculate the insulin to be delivered. A proportional-integral controller is used here for the calculation of the insulin to be delivered. The neural network was trained with data generated from a simulator called *AIDA* [55]. The algorithm was evaluated on five virtual patients, however, no specific results are presented in the paper.

Another approach is presented by Sanz et al. [82]. The approach is based on a so-called Disturbance Observer (DOB). The evaluation was done in two different models. In the first model, a closed-loop deployment of the algorithm is evaluated, which means that meal announcements are omitted as well. In the second model, meals are announced. The evaluation was performed using the UVa/Padova T1D simulator. The results for the closed-loop model are 80% Time-In-Range (TIR), while including meal announcements a Time-In-Range (TIR) of 88% was achieved.

A well-known approach by Atlas et al. [4] based on fuzzy logic consists of two control units, the Control to Range Module (CRM) and the Control to Target Module (CTM). The CRM is designed to maintain the glucose concentration in the patient's body within the range of 80-120 mg/dL, while the CTM is designed to maintain the concentration at a specified target value. The CRM receives the current glucose concentration, predictions of future glucose concentrations, and past and future trends in glucose concentration. The output of the CRM is a suggested change in basal rate as well as any insulin bolus that may be needed. These two values are then transferred to the CTM, which outputs the final insulin dose (divided into basal and bolus) with the inclusion of the Insulin-On-Board (IOB) and meal recognition. The authors evaluated their algorithm in a clinical setting on seven patients and achieved a Time-In-Range (TIR) of 73%. The Time-Above-Range (TAR) is presented as 27% while the Time-Below-Range (TBR) is 0%.

Another controller that is also based on fuzzy logic is known as the FL controller. It was developed by Mauseth et al [67, 68]. The control algorithm requires three inputs: current blood glucose (mg/dL), rate of change of blood glucose (mg/dL/min), and acceleration of blood glucose (mg/dL/min/min), while the output variable is the insulin microbolus dose for each point of time. The recommended insulin dose is then multiplied by the personalization factor to account for inter- and intrasubject variability. The algorithm was evaluated in a clinical setting with seven patients and achieved a Time-In-Range (TIR) of 76% over a 24-hour period. However, the Time-In-Range interval was defined here as 70-200 mg/dL.

Another approach is presented by Cobelli et al. [96] and is based on a model predictive control approach. The authors evaluate the algorithm on 100 simulated patients of the original UVa/Padova T1D simulator. They achieve a Time-In-Range (TIR) of 83.06%.

Another approach is presented by Hovorka et al. [43] and is also based on a model predictive control algorithm. The algorithm is evaluated in a clinical setting with ten patients over a period of 8-10 hours, with glucose concentration measured intravenously

every 15min. The authors obtained an average glucose concentration of 6.0 mmol/L measured (equivalent to 108 mg/dL).

Reinforcement learning is also quickly gaining popularity in the context of insulin delivery algorithms [94].

An approach by Zhu et al. [104] describes the use of Deep Reinforcement Learning based on Deep Q-Learning. This approach calculates a basal rate factor that is offset against a patient's current basal rate to keep them in the euglycemic range. In addition to this, meal boluses are delivered which are offset with variability. The results are evaluated with the UVa/Padova T1D simulator using ten adolescents and ten adult simulated patients. They use Time-In-Range (TIR), Time-Above-Range (TAR), Time-Below-Range (TBR), and Blood Glucose Risk Index (BGRI) as evaluation metrics. A Time-In-Range (TIR) of 80.94%±7.00 is obtained in the simulated patient base of ten adults and a Time-In-Range (TIR) of 65.85% ± 16.30 is obtained in the patient base of ten adolescents.

Another reinforcement learning approach is presented by Lee et al. [53]. This approach is based on an actor-critic model. Here an approach is presented that completely omits meal announcements and thus belongs to the closed-loop category. It was evaluated on the UVa/Padova T1D simulator, using ten adult and ten adolescent patients. Here, Lee et al. obtain an average glucose concentration of 124.72 ± 7.02 mg/dL and a Time-In-Range (TIR) of 89.56% ± 4.37.

In another approach by Fox et al. [37, 36] an algorithm based on Deep Q-Learning is developed. The algorithm is evaluated on the UVa/Padova T1D simulator. Unfortunately, however, the authors do not present concrete results with which a comparison could be made.

Another approach is presented by Yamagata et al. [103]. Here, a so-called Model-Based Reinforcement Learning approach is presented that combines a Model Predictive Controller approach with reinforcement learning. The approach is evaluated with the UVa/Padova T1D simulator utilizing three adults, three adolescents, and three children. They evaluate their approach using Time-In-Range (TIR). Here, they report the results of the nine patients evaluated individually. They achieve a TIR of 59.6% in child#1, a TIR of 55.3% in child#2, and a TIR of 45.1% in child#3. The adolescents achieve a TIR of 100.0% in adolescent#1, a TIR of n/a in adolescent#2, and a TIR of 66.1% in adolescent#3. Adults achieve a TIR of 56.8% in adult#1, a TIR of 73.3% in adult#2, and a TIR of 68.8% in adult#3.

Authors	Time-In-Range (%)	Method	Evaluation method
Sun et al. [90]	89.9 ± 8.7	Deep RL	UVa/Padova
Lee et al. [53]	89.56 ± 4.37	Deep RL	UVa/Padova
Sanz et al. [82]	80.0	DOB	UVa/Padova
Rossetti et al. [81]	78.8	PID	Clinical study
Atlas et al. [4]	73.0	Fuzzy logic	Clinical study

Table 3.1: Comparable approaches.

Another approach proposed is presented by Sun et al. [90]. This approach is based on an actor-critic approach (deep reinforcement learning). It is evaluated using the official UVa/Padova T1D simulator, using 100 simulated patients. They evaluate the quality of their algorithm primarily with the Time-In-Range (TIR), the Low Blood Glycaemic Index (LBGI) and the High Blood Glycaemic Index (HBGI). For this purpose, they simulate a period of three months. They use four different scenarios in which the TIR has been measured. The Time-In-Range obtained in the standard scenario is presented as $89.9\% \pm 8.7$.

Another approach, developed by Ngo et al. [70] is based on an actor-critic (deep reinforcement learning) approach. Unfortunately, no concrete results are shared in the paper, so it is not suitable for comparison.

An approach presented by El-Khatib et al. [33] describes an algorithm for closed-loop control using a bihormonal system. The authors evaluate their algorithm in a clinical setting on six patients over a 24-hour period, achieving an average glucose concentration of 140 mg/dL. Since this is a bihormonal system, it is not considered further in the comparison.

In Tab. 3.1 the results of some previous approaches are presented. These were selected based on whether evaluation results have been collected and whether they are comparable to each other. These approaches can only be compared to a limited extent, since all approaches have different experiment setups. In some cases, meal scenarios were implemented fixed, implemented with variation, or in some cases omitted. Evaluation parameters such as whether the UVa/Padova T1D simulator was used or if it was evaluated in a clinical setting also differ. The approaches listed here are most similar to the own experiment setup and therefore provide a basis for at least a rough comparison.

4 Data Resources

At the beginning of the research activity, a fundamental question to be answered is which data sources should be used for the training of the algorithm. Since reinforcement learning is a continuous agent system, the data source must be suitable for it. Suitability is composed of the fact that it either represents the metabolic state of a patient in real terms or has been shown to simulate it correctly. Availability also is an important consideration, as some data sources are not freely available. In this chapter, the data sources to be considered for the algorithm to be developed are elaborated. For this purpose, existing data from previous studies are first described. In the following, alternative data sources are discussed and the UVa/Padova T1D simulator is presented in detail.

4.1 OpenAPS Data Commons Dataset

In previous unpublished work, a dataset has already been obtained from the *OpenAPS Data Commons* program [56]. This data is subject to HIPAA privacy standards and is permitted to be used for the purpose of testing insulin delivery. The data includes N=82 records each containing several years of glucose concentration measurements, insulin deliveries, exercise activities, and therapy events such as carbohydrates added, infusion set changes etc. In a previous unpublished project, the data has already been successfully explored and first algorithms in the context of unannounced meal detection have been performed.

The OpenAPS dataset is in principle a suitable data source, but the algorithm requires a large number of primarily different metabolic situations that cannot be extracted from the dataset with sufficient diversity. In addition, implementing existing data into a reinforcement learning context would be of unpredictable complexity, which could potentially compromise the progress of the project. Therefore, the OpenAPS dataset is not considered further as a data source.

4.2 UVa/Padova T1D Simulator

In order to find a suitable alternative for the OpenAPS data, extensive research was conducted. The following two empirically collected data sources and two metabolism simulators widely used in numerous papers were found [100]:

1. PIMA dataset [22, 3, 40],
2. Ohio T1DM dataset [66, 91, 65, 30],
3. AIDA simulator [80, 52, 7],
4. UVa/Padova T1D Simulator [104, 90, 53].

The *PIMA dataset* is a dataset that originated from the *National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)*. It contains metabolic information from 768 patients, which allows predictions to be made as to whether a patient has diabetes or not. Only 268 patient records are classified as diabetic, while the other 500 patient records are classified as non-diabetic [69]. Even though there is metabolic data available from the 268 patients, as with the OpenAPS dataset, this dataset does not represent enough diverse metabolic situations and the data contain a too low resolution of information. Therefore, the data is not considered suitable for the intended use.

The *Ohio T1DM dataset* is a dataset created as part of the Blood Glucose Level Prediction (BGLP) Challenge of Ohio University. It contains continuous glucose data and other parameters such as meal intake and stress factors over eight weeks from 12 patients [64]. Because access to the dataset is limited to the scope of the Blood Glucose Level Prediction (BGLP) Challenge, the dataset is dropped from consideration.

The *AIDA simulator* is a freeware software developed by Lehmann et al. to simulate the metabolism of a diabetic patient. He points out that this simulator is to be used for educational purposes only and does not simulate accurately enough for research purposes [55]. Therefore, it is eliminated from consideration.

The *UVa/Padova T1D Simulator* [63] was developed by researchers of the University of Virginia and the University of Padova and is approved by the *American Food and Drug Administration (FDA)* to sufficiently simulate the metabolic system of type 1 diabetic patients. Previously, animal testing was used for this purpose. Several variable data sets can be generated based on different insulin levels, blood glucose levels and carbohydrate levels.

Since the UVa/Padova T1D simulator has official FDA approval compared to the AIDA simulator, it was decided to use the UVa/Padova T1D simulator for further development and the experiments to be performed.

A written request was made to use the official implementation of the UVa/Padova T1D simulator in the latest version S2013 for the purpose of this work. Unfortunately, the request was denied, which necessitated finding another implementation for this purpose.

In the course of further research, an open-source implementation of the UVa/Padova T1D simulator in version S2008 called *Simglucose* attracted attention. It is used in several research papers to simulate the patients metabolism [53, 77, 14, 98]. Simglucose [102] was implemented in Python and provides a reduced choice of 30 virtual patients (including ten adults, ten adolescents and ten children) in comparison to the original implementation. The respective patients all have individual metabolic parameters, which have been defined in Kovatchev et al. [49]. The differentiation into different age cohorts is based on the observed assumption that patients in different age groups exhibit certain metabolic peculiarities, such as children having a higher insulin sensitivity than adult patients [24]. These differences are taken into account by the simulator used.

Due to the limitation of the UVa/Padova T1D Simulator version S2008 to the simulation of insulin metabolism, only a mono-hormonal system will be developed and evaluated.

The different measurement frequencies of individual CGM systems from the interstitial glucose concentration are also taken into account by Simglucose. It is possible to determine the CGM system used for simulations. The following systems are available for selection:

1. Dexcom (3min.),
2. Guardian RT (5min.),
3. Navigator (1min.).

It is also possible to choose between two different insulin pump systems. The following systems are available for this purpose:

1. Insulet,
2. Cozmo.

The difference between the two systems are the limits of the maximum amount of insulin to be delivered in the form of a bolus or in the form of the basal rate. This must be taken into account when selecting the permitted actions of the algorithm, otherwise insulin may not be delivered when the algorithm actually intends to do so.

To illustrate how a simulation is performed in the UVa/Padova T1D simulator, two example plots generated by the simulator are briefly described.

In Fig. 4.1 is an exemplary daily simulation course shown, which demonstrates the use of the algorithm on the simulated patient *adult*#001 in the period of 24h (01 January to 02 January). The first graph from the top represents the blood and interstitial glucose progression over 24h, while the blood glucose concentration is represented by the blue line and the interstitial glucose concentration measured by the CGM is represented by the orange line. The background color shows the gradation of the individual reward limits. The green area represents the range of 90-110 mg/dL, the light red area displays the range of 110-140 mg/dL and the dark red area displays the range of 0-90 mg/dL and 140-401 mg/dL. The second graph from the top represents the amounts of carbohydrate supplied to the patient in grams per minute. The peaks show that punctual events took place here. The third graph from the top shows the amount of insulin delivered in units per minute, while the fourth graph from the top displays the patient's risk index in terms of the three previously described risk functions (Hypo Risk (LBGI), Hyper Risk (HBGI), Blood Glucose Risk Index (BGRI)). The fifth graph from the top represents the Insulin-On-Board (IOB) amount in units per minute. This figure indicates how much cumulative insulin is estimated to be still acting in the patient's body (see *Pharmacodynamics of insulin* in Background chapter).

In Fig. 4.2, which shows the simulation results of the experiments on several patients, the mean values of all measured parameters over a period of 24 h can be seen. In the first graph (from top to bottom), a dark blue line indicates the average measured glucose concentration in the blood. The bluish shaded area indicates the standard deviation of $\pm$ 1. The green dashed line indicates the lower limit of the euglycemic range (70 mg/dL), while the red dashed line represents the upper limit of the euglycemic range (180 mg/dL). The second graph from the top shows the average measured interstitial fluid glucose level. This differs from the blood glucose level because the simulator includes a manufacturer-related measurement error. The third graph (from top to bottom) shows the average amount of carbohydrate consumed in grams per minute. This is colored in a blue tone and scatters around the predefined meal scenario.

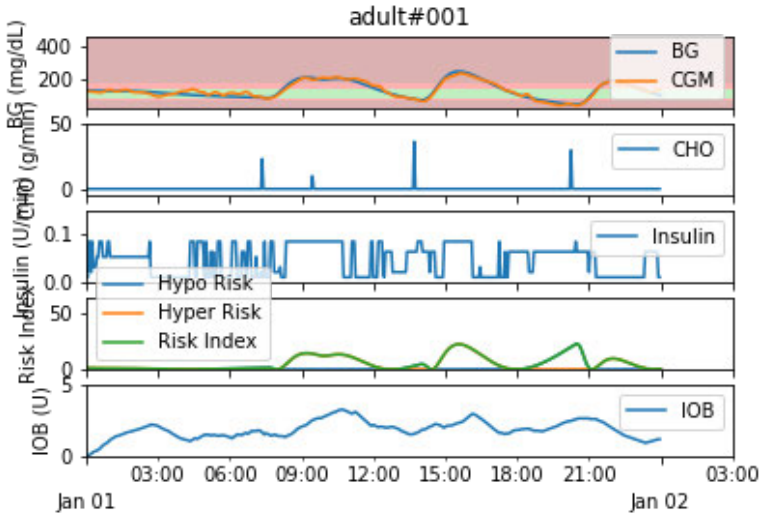

Figure 4.1: Exemplary daily course of metabolic parameters of a simulated patient.

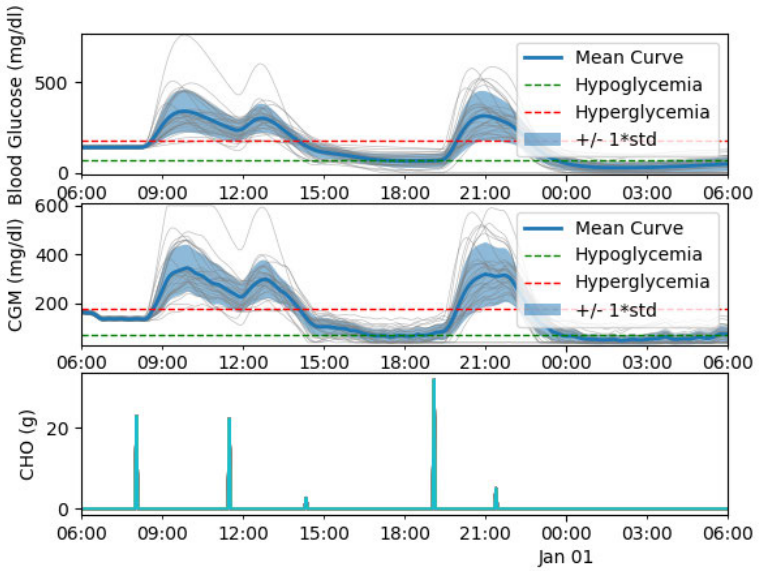

Figure 4.2: Exemplary simulation course (mean values over several simulated patients).

5 Experiment Setup

In order to answer the proposed research questions adequately, experiments will be conducted. For this purpose, this chapter describes the experimental setup for performing the experiments. At first, the implemented Deep Q-Learning algorithm is described. Then, in the second section, the meal scenario to be used is presented. In the next section, the parameters used in the experiments are described. After that, the evaluation metrics used are presented. In the last section it is briefly described which extensions had to be implemented in the Simglucose simulator and which logging tool and logged parameters were used for the experiment observation during the training process of the underlying neural network.

5.1 Implementation details

It is necessary to first describe a few implementation details in order to fully understand the experiments to be conducted.

The chosen UVa/Padova T1D simulator generates a glucose measurement every three minutes, which is called observation o_t in this context. In general after the observation o_t is received, an input vector will be prepared with additional information and an action a_t is executed. The measurement frequency of three minutes was chosen because most modern commercial CGM systems output a reading at this frequency. Furthermore, this time period is common in current Artificial Pancreas systems [48].

The agent receives a multidimensional input vector $D = \{G; M\} = [d_{t+1-L}, ..., d_t] \in \mathbb{R}^{1\times L}$ which consists of the continuously measured glucose concentration (G) and the supplied meals (M). The length L defines how long the measurement sequences are chosen to be used for the algorithm in the training procedure and simulation procedure (see Tab. 5.1).

The algorithm to be developed should also handle ingested meals without delivered meal boluses so no bolus calculator was used for the experiments. Thus, the experiment results and the proposed algorithms are considered as closed-loop approaches.

If the measured glucose concentration is outside the interval of 30 mg/dL and 350 mg/dL, the training run is aborted and restarted.

The algorithms training process is described in its essential steps in Alg. 1.

Algorithm 1 Implementation of Deep Q-Learning.

Input: the target network update period T_{main}, exploration before learning exp, training period T_{target}, training episodes E_{train}.
Initialize DQNs with random weights θ_1,θ_2, replay memory B
for step $t \in 1, 2, ...exp$ **do**
    Choose action randomly, observe o', obtain r, store (o, a, r, o') into B
end for
for step $t \in 1, 2, ...E_{train}$ **do**
    Sample action ϵ-greedy from DQN, observe o', obtain r, store (o, a, r, o') into B
    if t mod T_{target} **then**
        Sample a mini-batch uniformly from B and calculate loss
        Perform a gradient descent to update θ_1
    end if
    if t mod T_{main} **then**
        $\theta_2 \leftarrow \theta_1$
    end if
end for

For the underlying units of the Deep Q-Networks, *Long Short-Term Memory (LSTM)* units were chosen [41]. These are able to recognize and learn even long-term correlations from the given input data. For the optimization of the learning rate the Adam optimizer is used [47]. Furthermore, the Deep Q-Networks get trained with batches of size b as shown in Tab. 5.1.

5.2 Meal scenario

To answer the research question of to what extent the algorithm can compensate for a patient's meals with the basal rate, a meal scenario must first be created. This scenario should represent deviations in regular eating behavior in order to adequately represent a patient's daily routine.

For this purpose, a meal scenario was chosen that provides a pattern of four meals per day with varying meal sizes:

1. 7 am (70g),
2. 10 am (30g),
3. 2 pm (110g),
4. 9 pm (90g).

Times were set with a variability of one standard deviation of 60min. and a meal size variability with a coefficient of variation of 10% to reflect the variation in eating behavior through a patient's daily life [104].

5.3 Parameter

For the first experiment, the hyper-parameters shown in Tab. 5.1 were used following the previously evaluated parameters of Zhu et al. [104]. They serve as the basis for all experiments, but are also adapted depending on the experiment.

Parameter	Value
Exploration before learning	128
Target network update period T_{target}	100
Update greedy ϵ	$0.5 \rightarrow 0.01$
Replay memory size	50000
Adam learning rate	1×10^{-3}
Number of timesteps L	1
Training episodes E_{train}	1000
Discount factor γ	0.99
Number of layers	3
Dropout between layers	0.2
Number of hidden units	64
Batch size b	64

Table 5.1: List of hyperparameters used in the conducted experiments.

The following allowed actions were defined for the experiments:

$A = [0 * BR, 0.5 * BR, 1 * BR, 1.5 * BR, 2 * BR, 3 * BR, 4 * BR]$.

The basal rate BR is defined differently for each simulated patient and is already known to the algorithm in advance. The actions were selected up to 400% of basal rate to be able to handle larger amounts of added food.

To avoid unexpected rejections of the actions during training and evaluation of the algorithm, the Cozmo pump is selected for both procedures. This allows a maximum basal rate delivery of 35 U/hr compared to the Insulet pump, which allows a maximum basal rate delivery of only 30 U/hr.

For the experiments, the reward function previously evaluated by Zhu et al. (see Eq. 5.1) was used [104]. It defines the reward r at time step t section-wise, depending on the measured glucose concentration G at the next time step $t+1$.

$$r_t = \begin{cases} 1, & 90 \leq G_{t+1} \leq 140 \\ 0.1, & 70 \leq G_{t+1} < 90 \;\&\; 140 < G_{t+1} \leq 180 \\ -0.4 - (G_{t+1} - 180)/200, & 180 < G_{t+1} \leq 300 \\ -0.6 + (G_{t+1} - 70)/100, & 30 \leq G_{t+1} < 70 \\ -1, & else. \end{cases} \tag{5.1}$$

The algorithm receives a positive reward when it is in the euglycemic range and negative rewards when it is outside of this range. A hypoglycemic glucose concentration is penalized more severely than a hyperglycemic glucose concentration in this function because hypoglycemias pose a more severe threat to metabolism than hyperglycemias, and these can only be responded to with manual meal delivery [104].

As the list of used hyperparameters (see Tab. 5.1) show, the experiments were conducted with a training phase of 1000 episodes, since the loss of the neural network converges against 0 at that number of episodes.

For reproducibility of results, all random-producing code components were initialized with a random seed of 42.

5.4 Evaluation metrics

In order to evaluate the developed algorithm according to the proposed research questions, relevant evaluation metrics must first be selected.

The following metrics were selected based on Danne et al., Battelino et al., and Maahs et al. [29, 9, 61]:

1. Time-In-Range (TIR),
2. Time-Below-Range (TBR),
3. Time-Above-Range (TAR),
4. Low Blood Glucose Risk Index (LBGI),
5. High Blood Glucose Risk Index (HBGI),
6. Blood Glucose Risk Index (BGRI).

As previously described the HbA_{1C} is an indicator that can only describe if the long-term therapy needs of a patient are met. It is therefore more suitable to evaluate the long-term effect of the proposed insulin delivery algorithm in a clinical trial setting, but not the simulation of a diabetic patient. It cannot represent acute glycemic abnormalities such as hypoglycemia or hyperglycemia [9].

5.5 Simglucose extensions

In the course of the project, some changes and improvements also had to be made to the implementation of the simulator. For this purpose, the simulator's repository was loaded directly into the project in order to modify it according to the project's needs.

For some experiments additional simulation parameters were needed, which were missing in the original implementation. This refers especially to the implementation of the Insulin-On-Board (IOB) calculation.

5.6 Logging

In order to be able to observe the progress of the training of the proposed Deep Q-Learning algorithm, the tool *Weights&Biases* was chosen [15]. This tool allows comparatively simple logging of the most relevant parameters, logs of the achieved rewards, Time-In-Range (TIR) and the agents performed actions.

The Time-In-Range (TIR) did not provide enough information about a run, since it depends on the number of timesteps achieved. It can happen that a TIR of 90% is achieved but only one hour was simulated (i.e. low timesteps are logged). For this reason, the measured Time-In-Range (TIR) was divided by the number of achieved timesteps of a run. Another logged metric is the *Quality*, which is calculated by dividing the achieved rewards of an episode by the achieved timesteps. It provided additional information about the success of the conducted training run.

The average score over the most recent 100 episodes was also logged to monitor and compare the trend during the training process.

6 Experiments

Following the experiment setup, this chapter describes the conducted experiments in detail and presents them in tabular and graphical form. The results of the experiments are evaluated and summarized in tabular form after the detailed description.

For each of the experiments, two additional plots were created to show the different glucose ranges per patient and showing the risk assessment per patient represented in a bar chart. Since these plots only differ in their type of representation from the plots shown, they have been moved to the appendix. The results in the presented tables were rounded to two decimal places.

6.1 PID controller

A simple PID controller was first implemented in order to be able to compare the algorithm to be developed with at least one fully matching experimental setup.

It operates according to the scheme shown in Alg. 2.

Algorithm 2 Implementation of a simple PID Controller.

1: Input: the glucose concentration G_{now} provided by the simulator, the target glucose concentration G_{target}, the tuned P, I and D factors.
2: **repeat**
3: Obtain glucose concentration and store it in G_{now}.
4: $action = P * (G_{now} - G_{target}) + I * integratedState + D * (G_{now} - previousBG)$
5: $previousBG = G_{now}$
6: $integratedState = integratedState + (G_{now} - G_{target})$
7: **until** simulation over

The implemented controller was evaluated with the experiment setup presented in the previous chapter, using the parameters shown in Tab. 6.1.

Parameter	Value
Tuning factor P	0.0001
Tuning factor I	0.00001
Tuning factor D	0.001
Target glucose G_{target}	120

Table 6.1: List of PID parameters.

As the simulation results (see Tab. 6.2) show, the mean Time-In-Range (TIR) across all simulated adult patients is $46.84\% \pm 13.29$, while the mean Time-Above-Range (TAR) is $15.8\% \pm 5.50$ and the mean Time-Below-Range (TBR) is $37.36\% \pm 14.15$.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	51.98	8.32	39.71	37.19	1.68	38.87
adult#002	68.61	13.51	17.88	2.26	1.2	4.26
adult#003	40.75	21.21	38.05	20.82	7.03	27.85
adult#004	21.62	13.51	64.86	50.09	3.55	53.63
adult#005	40.75	8.73	50.52	20.91	2.29	23.2
adult#006	40.12	23.70	36.17	61.91	5.84	67.75
adult#007	40.96	12.47	46.57	16.13	3.35	19.47
adult#008	62.58	14.76	22.66	45.03	3.51	48.55
adult#009	46.99	20.58	32.43	13.45	3.82	17.27
adult#010	54.05	21.21	24.74	8.83	2.68	11.52

Table 6.2: Simulation performance of PID controller for adult population.

The glucose curve (see Fig. 6.1) shows, that the PID controller in the preprandial phase (in the first six hours) is in the euglycemic range with a slight tendency towards hypoglycemia. On the basis of the glucose concentration peaks, it can be deduced that the PID controller responds to the first meal consumed with a slightly too high amount of insulin and then delivers too much insulin for the second and third meal, so that the patients experience severe hypoglycemia.

The CVGA (see Fig. 6.2) shows that the risk of 70% of all simulated patients is located in one of the D fields and of 30% in the E field. It can be deduced that the extreme glycemic deviation is in both the hyperglycemic and hypoglycemic directions.

This could result from the fact that the controller only responds directly to glucose concentrations and has no knowledge of the amount of residual insulin already in the body that is still acting (IOB). The problems could also possibly be solved by optimizing the only roughly tuned PID controller variables reasonably, but within the scope of

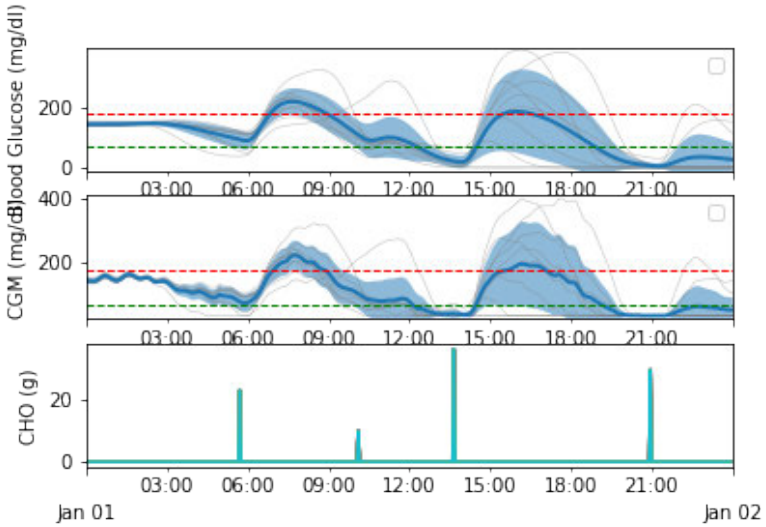

Figure 6.1: Simulation plot of the PID controller (mean values over adult population).

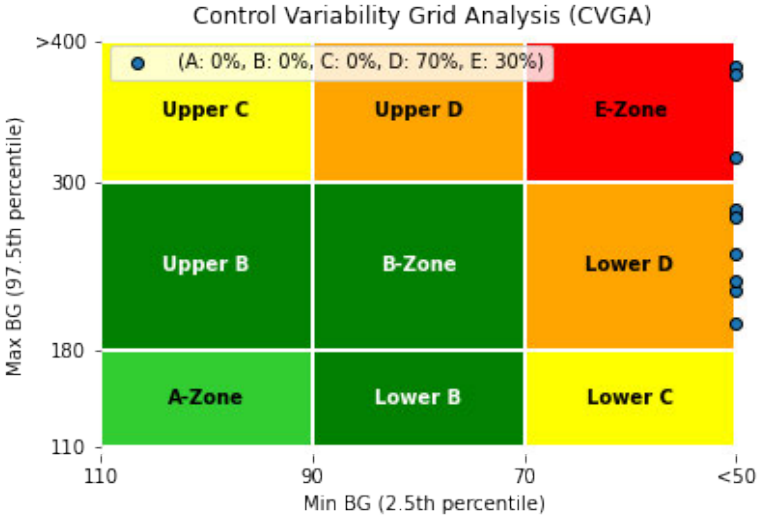

Figure 6.2: CVGA plot of the PID controller.

this thesis this is not in focus, since the PID controller is only intended to serve as a comparative alternative.

6.2 Baseline implementation

Now that the implemented PID controller has been evaluated, a first experiment with the Deep Q-Learning implementation will be evaluated.

For this implementation, different configurations of the size of the neural network used were tried. In order to evaluate which configuration is the most successful, the average score of the last 100 episodes within the training phase is considered as a first indicator of success. The results shown in Fig. 6.3 indicate that the 64 units x 3 layer configuration is the most successful configuration in the last episode. Therefore, this configuration will be used as a basis for the following experiments.

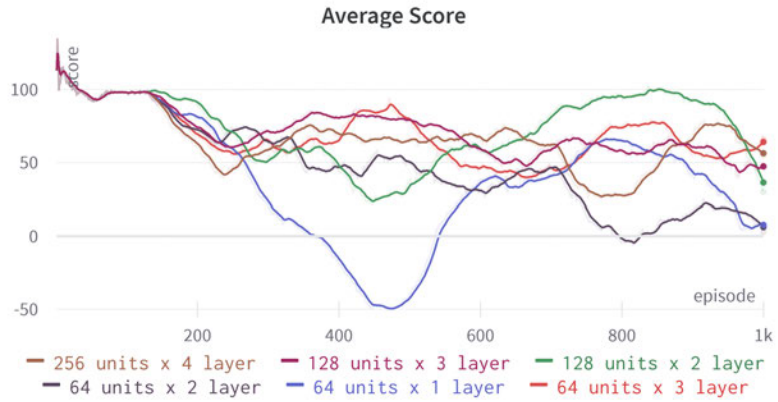

Figure 6.3: Comparison of different neural network configurations.

As the simulation results (see Tab. 6.3) show, the mean Time-In-Range (TIR) across all simulated adult patients is $52.24\% \pm 15.81$, while the mean Time-Above-Range (TAR) is $46.22\% \pm 18.37$ and the mean Time-Below-Range (TBR) is $1.54\% \pm 4.86$.

The glucose curve is shown in Fig. 6.4. The results shown in Tab. 6.3 show that the number of hypoglycemic episodes is very low. Only the patient *adult*#007 is 15.38% below the euglycemic range. Furthermore, it shows that the span of the Time-In-Range (TIR) is very high. The worst simulated patient has a TIR of 27.44%, while the best simulated patient experiences a TIR of 70.89%.

As shown in the CVGA (see Fig. 6.5) the variability of the glucose concentration compared to the previously evaluated PID controller slightly improved for a few patients.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	53.43	46.57	0.00	0.0	10.74	10.74
adult#002	65.90	34.10	0.00	0.0	7.75	7.75
adult#003	57.17	42.83	0.00	0.002	12.96	12.97
adult#004	54.26	45.74	0.00	0.04	14.34	14.34
adult#005	36.59	63.41	0.00	0.0	18.17	18.17
adult#006	56.55	43.45	0.00	0.09	13.16	13.16
adult#007	70.89	13.72	15.38	1.49	1.81	1.81
adult#008	70.06	29.94	0.00	0.0	5.40	5.40
adult#009	27.44	72.56	0.00	0.0	27.94	27.94
adult#010	30.15	69.90	0.00	0.0	25.78	25.78

Table 6.3: Simulation performance of 64 units x 3 layer configuration for adult population.

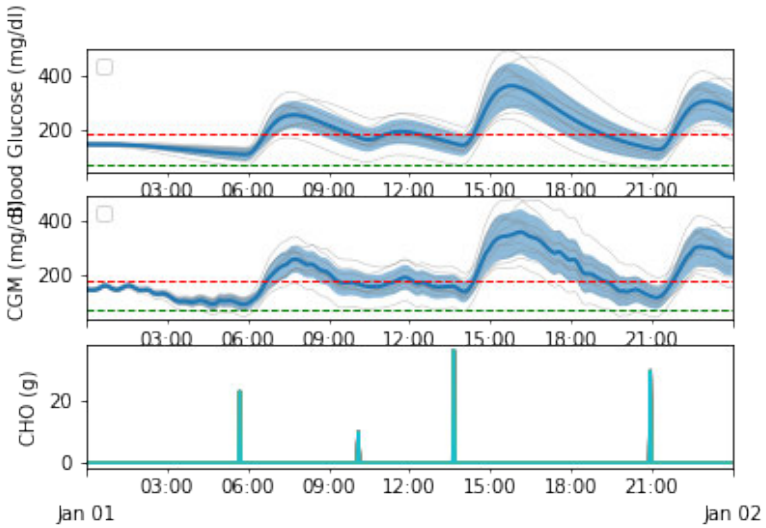

Figure 6.4: Simulation plot (mean values over adult population) of baseline implementation.

Only 20% of all simulated patients are in one of the D fields, while another 20% are in the Upper-C field. An additional 10% of simulated patients are in the Upper-B field.

The high span of the TIR results could result from the fact that the algorithm has only been trained on the patient *adult*#001 and thus the specific metabolic parameters of this patient have been learned. It is possible that the range behaves similarly across all further experiments. However, perhaps a stronger generalization can be obtained. Thus, it will be important to observe how the span behaves in the further experiments.

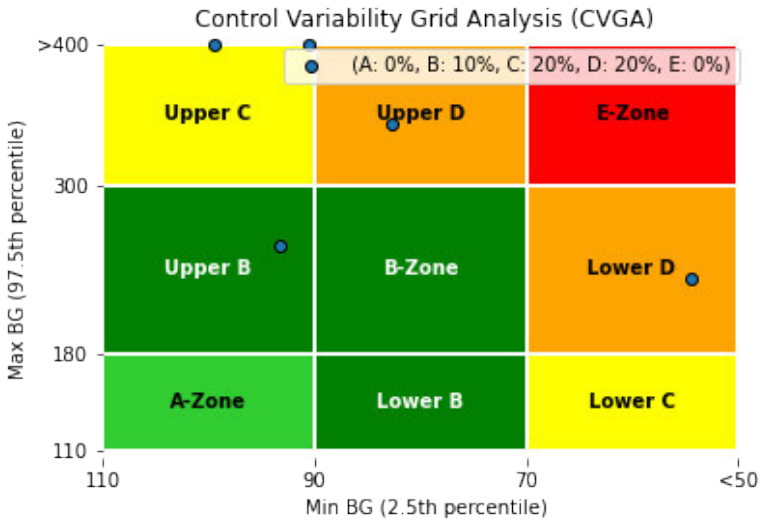

Figure 6.5: CVGA plot of baseline implementation.

6.3 Apply scaling

Since it generally has many advantages to scale the input data to a range between 0 and 1, an experiment will be performed in which the input data is scaled that way. For the process of scaling, the *Scikit-learn* MinMax-Scaling of the form as shown in Eq. 6.1 was chosen.

$$\begin{aligned} X_{std} &= (X - X.min(axis = 0))/(X.max(axis = 0) - X.min(axis = 0)) \\ X_{scaled} &= X_{std} \times (max - min) + min \end{aligned} \tag{6.1}$$

The use of this scaling is based on the assumption of robustness to very small standard deviations of features and preservation of zero entries in sparse data [74].

As the simulation results (see Tab. 6.4) show, the mean Time-In-Range (TIR) across all simulated adult patients is $50.62\% \pm 12.90$, while the mean Time-Above-Range (TAR) is $15.55\% \pm 11.32$ and the mean Time-Below-Range (TBR) is $33.83\% \pm 19.76$.

The glucose concentration (see Fig. 6.6) tends to mild hypoglycemia already in the preprandial phase (first six hours). The first meal at just before 6 am causes glucose concentration to rise on average slightly above 180 mg/dL. The second meal allows the

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	62.79	9.56	27.65	3.73	1.08	4.81
adult#002	64.24	4.99	30.77	4.29	0.73	5.01
adult#003	38.88	21.0	40.12	20.47	1.87	22.34
adult#004	36.8	22.66	40.54	11.11	2.12	13.23
adult#005	62.99	23.70	13.31	1.27	4.02	5.28
adult#006	39.92	10.60	49.48	18.34	1.57	19.91
adult#007	34.72	2.70	62.58	32.58	0.0	32.58
adult#008	42.41	0.0	57.59	32.49	0.18	32.67
adult#009	61.75	31.19	7.07	0.17	6.64	6.82
adult#010	61.75	29.11	9.15	0.9	6.75	7.65

Table 6.4: Performance statistics of applied scaling experiment for adult population.

glucose concentration to remain in the euglycemic range. At the last two meals, the glucose concentration rises to a mean of about 200-250 mg/dL. The postprandial phase of the third meal takes place in the relatively strong hypoglycemic range.

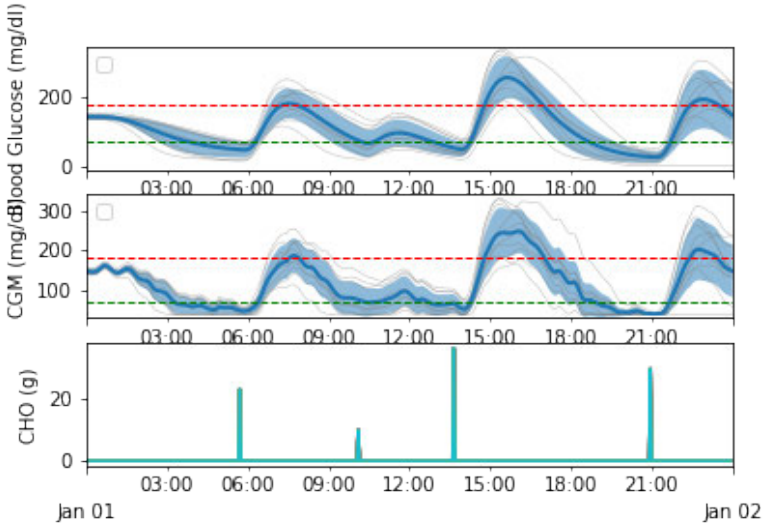

Figure 6.6: Simulation plot (mean values over adult population) of applied scaling experiment.

The CVGA (see Fig. 6.7) shows that 50% of all simulated patients are located in one of the D fields and 40% even in the E field. Another 10% of the patients are in the Upper-C field.

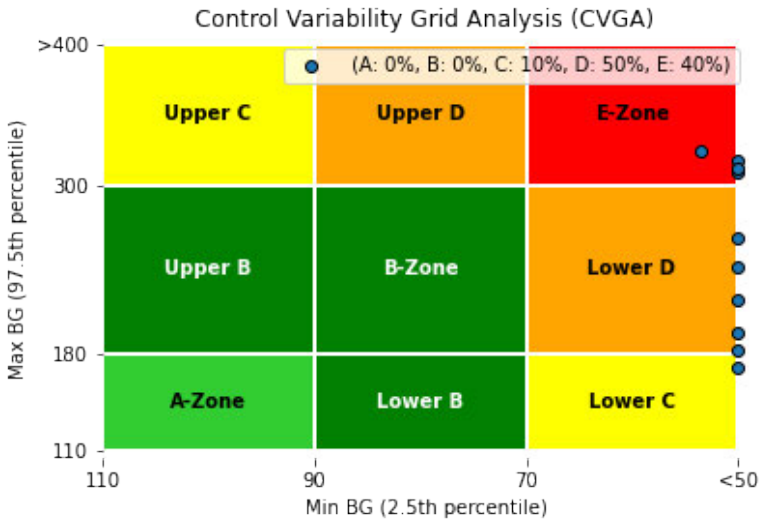

Figure 6.7: CVGA plot of applied scaling experiment.

Compared to the evaluation of the previous implementation, the ingested meals are intercepted much better, which means that the peak glucose concentrations are not as high as in the previous approach. However, it can also be observed that the simulated patients tend to have severe hypoglycemia (BG < 50) in the postprandial phases. This can also be seen in Fig. A.6, where severe hypoglycemia accounts for the majority of hypoglycemias in most patients.

This could result from the fact that the algorithm does not yet know how much insulin is currently acting in the body. Including the Insulin-On-Board (IOB) as part of the input vector could improve this situation. Whether this is the case will be investigated in a follow-up experiment.

It also might be possible that the algorithm has already recognized the regularity of the meal scenario and already increases the insulin supply before the presumed start of a meal. In this way, it would tolerate hypoglycemia in order to keep the subsequent glucose peaks low. Further observation is required to determine whether this effect continues to occur and whether it can be avoided to prevent hypoglycemias.

6.4 Using Insulin-On-Board (IOB)

A few research papers report that using Insulin-On-Board (IOB) as a factor to be considered yields better results in general obtained TIR [81]. Therefore, the following experiments will evaluate the impact of IOB as an input factor for the presented algorithm.

For this purpose, a first simple modeling of the Insulin-On-Board is used, which is bilinear in nature. In a follow-up experiment, the IOB calculation with exponential decay is then evaluated. Both modelings were implemented according to the implementation of *OpenAPS* [97]. The variable DIA (duration of insulin activity) specifies the maximum duration of action of the insulin to be modeled in hours, while the peak specifies the maximum time of action of the insulin since delivery of the insulin in minutes. The bilinear model (see Fig. 6.8) can be perceived to consist of two simple linear functions, the first of which peaks at the assumed maximum effect of the insulin used, and the second of which peaks at the maximum duration of action of the insulin from the maximum effect.

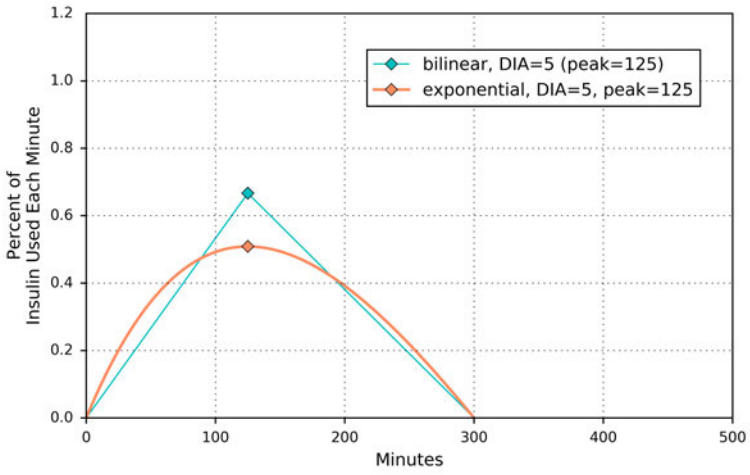

Figure 6.8: Bilinear vs. exponential IOB calculation [97].

The exponential model looks more like the shape of a downward opening parabola. This corresponds more closely to the pharmacodynamics of fast-acting commercially available insulins and therefore promises better simulation results. The amount of acting insulin grows faster in the exponential model and flattens out differently than in the bilinear model. Due to the different distribution, the amount of insulin acting at peak time is

lower in the exponential model than in the bilinear model. However, the total amount of acting insulin is the same for both models.

6.4.1 Bilinear model

In a first experiment, modeling of the IOB based on the bilinear model is used. This allows a comparatively simple calculation of the Insulin-On-Board (IOB) (see Fig. 6.8).

As the simulation results (see Tab. 6.5) show, the mean Time-In-Range (TIR) across all simulated adult patients is $48.02\% \pm 27.95$, while the mean Time-Above-Range (TAR) is $48.77\% \pm 31.79$ and the mean Time-Below-Range (TBR) is $3.20\% \pm 6.98$.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	49.9	50.10	0.0	0.0	24.29	24.29
adult#002	82.74	17.26	0.0	0.0	4.22	4.22
adult#003	25.78	74.22	0.0	0.0	104.52	104.52
adult#004	11.23	88.77	0.0	0.0	115.74	115.74
adult#005	35.14	64.86	0.0	0.0	30.8	30.8
adult#006	70.69	18.92	10.4	0.23	4.04	4.27
adult#007	74.43	4.78	20.79	2.83	1.04	3.88
adult#008	84.2	14.97	0.83	0.15	2.83	2.98
adult#009	27.65	72.35	0.0	0.0	36.7	36.7
adult#010	18.50	81.5	0.0	0.0	69.75	69.75

Table 6.5: Performance statistics of bilinear IOB modeling experiment for adult population.

The glucose progression is shown in Fig. 6.9. The results are noticeably worse compared to the simulation without including IOB as a factor. Especially the meals are not answered correctly and let the average glucose concentration of the simulated patients rise far into the hyperglycemic range (up to 600 mg/dL). Furthermore, the standard deviation is very high, suggesting that a good generalization of the algorithm has not been achieved.

In the CVGA (see Fig. 6.10), it shows that 20% of all simulated patients are in one of the D fields, while another 20% are in one of the B fields.

This progression into the hyperglycemic range could be due, in part, to the simplicity of the bilinear modeling of IOB. Therefore, a follow-up experiment will evaluate the use of exponential IOB modeling.

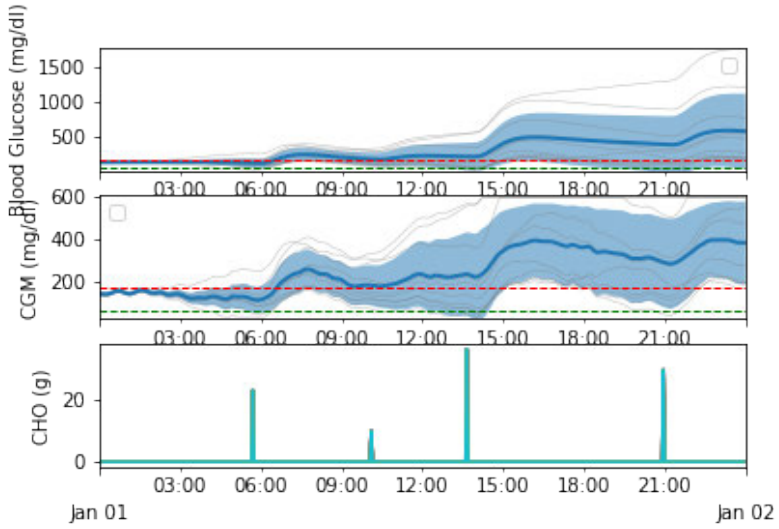

Figure 6.9: Simulation plot (mean values over adult population) of bilinear IOB experiment.

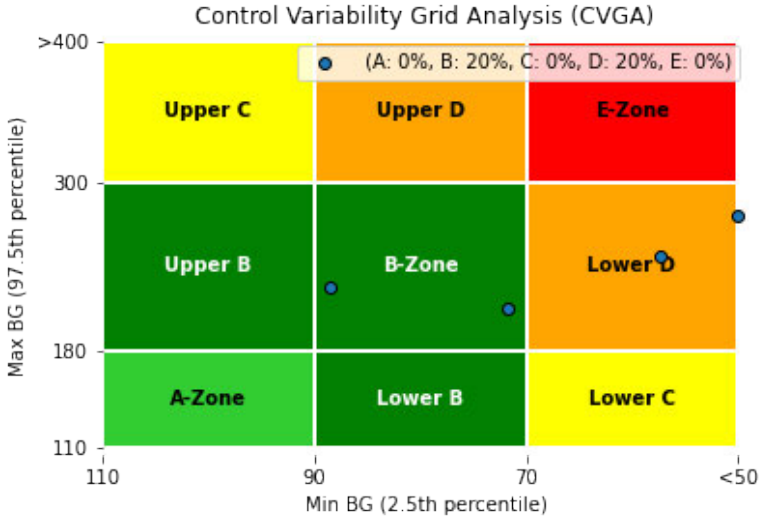

Figure 6.10: CVGA plot of bilinear IOB experiment.

6.4.2 Exponential model

Since the initial modeling of the Insulin-On-Board (IOB) did not yield an improvement in the control strategy, a more detailed modeled variant is now evaluated. For this purpose, the IOB is now implemented with an exponential model (see Fig. 6.8). Since this model

appears to provide a much more accurate representation of the pharmacodynamics of insulin, a significant improvement in simulation results is suspected.

Before conducting the experiment the exponential model of the IOB requires two parameters. One is the peak, which is the assumed time of maximum effect of insulin in the body, and the DIA, which is the assumed maximum duration of effect of insulin in the body. A first experiment is performed with a DIA of three hours and a peak of 75 minutes.

As the simulation results (see Tab. 6.6) show, the mean Time-In-Range (TIR) across all simulated adult patients is $57.71\% \pm 12.12$, while the mean Time-Above-Range (TAR) is $23.02\% \pm 14.15$ and the mean Time-Below-Range (TBR) is $19.27\% \pm 18.08$.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	70.06	19.13	10.81	0.76	2.32	3.08
adult#002	74.43	14.97	10.60	1.09	2.01	3.1
adult#003	44.49	24.32	31.19	10.03	2.66	12.69
adult#004	44.7	22.45	32.85	6.05	2.34	8.38
adult#005	66.53	33.47	0.0	0.07	8.4	8.47
adult#006	47.82	14.35	37.84	7.34	2.25	9.59
adult#007	46.36	2.29	51.35	12.95	0.0	12.95
adult#008	72.56	9.98	17.46	4.41	1.10	5.51
adult#009	53.22	46.15	0.62	0.01	11.74	11.75
adult#010	56.96	43.04	0.0	0.05	12.32	12.37

Table 6.6: Performance statistics of exponential IOB experiment for adult population.

The glucose curve (see Fig. 6.11) shows that the simulated patients are prone to hypoglycemia in the preprandial phase (first six hours) but are on average within the euglycemic range at the lower edge (80 mg/dL). The first meal results in a peak glucose concentration just above 200 mg/dL. After the first meal, patients fall into the euglycemic range on average. The second meal barely changes this situation. The third meal causes patients to rise to a mean glucose concentration of slightly less than 300 mg/dL. Postprandially, patients experience mild hypoglycemia on average after the third meal. The fourth meal causes the glucose concentration to rise to a mean of slightly less than 250 mg/dL.

The CVGA (see Fig. 6.12) shows that 70% of the simulated patients are located in one of the D fields, 20% in the E field, and another 10% in the lower-C field.

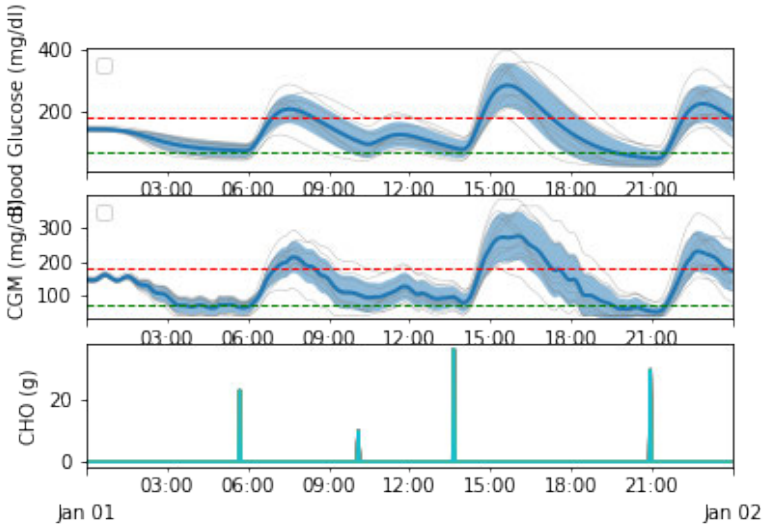

Figure 6.11: Simulation plot (mean values over adult population) of exponential IOB experiment.

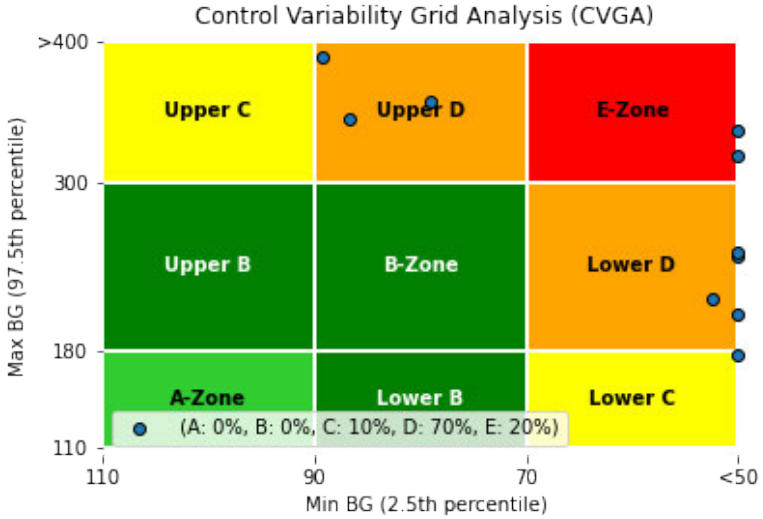

Figure 6.12: CVGA plot of exponential IOB experiment.

The simulation results (see Tab. 6.6) show that the use of exponential modeling of the Insulin-On-Board (IOB) achieves a significant improvement in the simulation results. Compared with the configuration without the inclusion of Insulin-On-Board (IOB) as a factor, the mean TIR (%) improved by 7.09. The standard deviation has decreased by 0.78. Also, compared to the configuration with bilinear modeling of Insulin-On-Board

(IOB), the average TIR (%) has improved by 9.69. The standard deviation has decreased by 15.83.

However, it can also be observed that the TBR is still relatively high. Especially in the later course of the simulation, the patients experience hypoglycemia on average. In Fig. A.10 it can be seen that the time patients experience severe hypoglycemia (BG < 50) clearly outweighs mild hypoglycemia in patients *adult*#003, *adult*#004 and *adult*#007. This could be due to an incorrectly assumed DIA of three hours. As a result, too little IOB could be assumed in the later hours, which leads to a higher amount of insulin being delivered than necessary and the patient reaching hypoglycemia. However, it could also indicate that the action space is not selected in a refined enough way. Using the current action space the algorithm must react to the taken meals starting from 200% in 100% steps, which may be a too coarse gradation.

To investigate this assumption, a follow-up experiment with a DIA of five hours and a peak of 75 minutes will be performed. It shall be observed whether the longer assumed insulin action time lowers the TBR in general and especially in the later simulation course. Another experiment will investigate whether a more refined action space can prevent postprandial hypoglycemia.

6.4.3 Increase DIA from 3h to 5h

To evaluate whether the postprandial hypoglycemias from the previous experiment can be avoided by using a differently chosen Duration of Insulin Activity (DIA), an experiment with a DIA of five hours instead of the previous three hours will now be conducted.

As the simulation results (see Tab. 6.7) show, the mean Time-In-Range (TIR) across all simulated adult patients is $40.35\% \pm 13.04$, while the mean Time-Above-Range (TAR) is $53.62\% \pm 23.33$ and the mean Time-Below-Range (TBR) is $6.03\% \pm 16.99$.

The glucose progression (see Fig. 6.13) shows that the preprandial phase (first six hours) is optimally maintained in the euglycemic range. The first meal ingested results in a peak glucose concentration averaging 250 mg/dL. The postprandial phase of the first meal occurs at the upper end of the euglycemic range. The second meal also causes the glucose concentration to rise only briefly and slightly above the euglycemic range on average. The third meal results in a peak glucose concentration averaging 400 mg/dL.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	38.46	61.54	0.0	0.0	15.57	15.57
adult#002	28.27	71.73	0.0	0.0	21.85	21.85
adult#003	49.9	50.10	0.0	0.02	23.36	23.37
adult#004	34.93	65.07	0.0	0.16	68.31	68.47
adult#005	32.64	67.36	0.0	0.0	24.14	24.14
adult#006	53.85	39.92	6.24	0.65	20.15	20.8
adult#007	45.95	0.0	54.05	15.88	0.01	15.89
adult#008	65.49	34.51	0.0	0.0	8.33	8.33
adult#009	27.44	72.56	0.0	0.0	37.28	37.28
adult#010	26.61	73.39	0.0	0.0	47.17	47.17

Table 6.7: Performance statistics of exponential IOB (DIA 5h) experiment for adult population.

The postprandial phase of this meal is permanently above the euglycemic range from this point on.

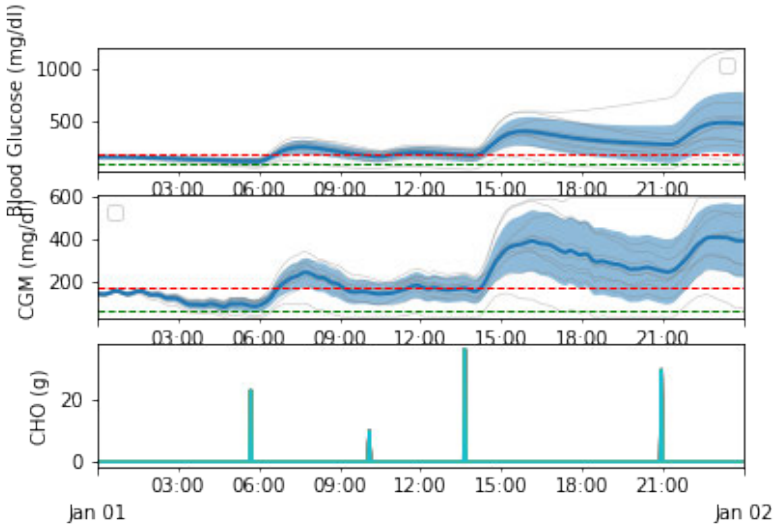

Figure 6.13: Simulation plot (mean values over adult population) of exponential IOB (DIA 5h) experiment.

The CVGA (see Fig. 6.14) shows that 20% of the simulated patients are located in one of the C fields, while another 10% are located in the E field. Another 10% of the patients are in the Upper-D field.

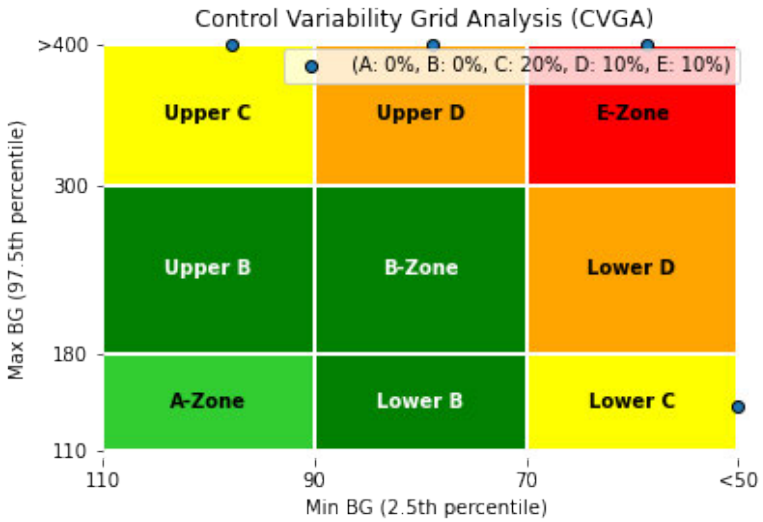

Figure 6.14: CVGA plot of exponential IOB (DIA 5h) experiment.

As can be seen in Fig. 6.13, the algorithm does not seem to perform reasonable insulin deliveries. Although mild and severe hypoglycemia are avoided, prolonged glucose concentrations above 300 mg/dL imply a high risk for patients and cannot be tolerated. Thus, increasing the DIA from three hours to five hours did not show any improvements in the simulation results and will not be adapted for the upcoming experiments.

The next experiment aims to evaluate whether the use of a more refined action space yields better results in TIR and reduces the Time-Below-Range (TBR).

6.5 Using refined action space

So far, in all experiments performed, an action space of the form

$$A = [0 * BR, 0.5 * BR, 1 * BR, 1.5 * BR, 2 * BR, 3 * BR, 4 * BR]$$

was used.

To investigate whether the postprandial hypoglycemias of the simulated patients can be prevented with a more refined action space and improve the overall TIR, we use an action space that operates in steps of 25% for the following experiment. This results in an action space of:

$A = [0 * BR, 0.25 * BR, 0.5 * BR, 0.75 * BR, 1 * BR, 1.25 * BR, 1.5 * BR, 1.75 * BR, 2 * BR, 2.25 * BR, 2.5 * BR, 2.75 * BR, 3 * BR, 3.25 * BR, 3.5 * BR, 3.75 * BR, 4 * BR]$.

As the simulation results in Tab. 6.8 show, the mean Time-In-Range (TIR) across all simulated adult patients is $27.09\% \pm 1.06$, while the mean Time-Above-Range (TAR) is $72.91\% \pm 1.06$ and the mean Time-Below-Range (TBR) is $0.00\% \pm 0.00$.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	27.03	72.97	0.0	0.0	23.14	23.14
adult#002	28.07	71.93	0.0	0.0	18.22	18.22
adult#003	26.2	73.80	0.0	0.0	33.85	33.85
adult#004	25.99	74.01	0.0	0.0	48.65	48.65
adult#005	27.03	72.97	0.0	0.0	26.67	26.67
adult#006	27.23	72.77	0.0	0.0	48.72	48.72
adult#007	27.03	72.97	0.0	0.0	32.61	32.61
adult#008	29.52	70.48	0.0	0.0	17.51	17.51
adult#009	26.82	73.18	0.0	0.0	38.39	38.39
adult#010	25.99	74.01	0.0	0.0	38.03	38.03

Table 6.8: Performance statistics of refined action space experiment for adult population.

The glucose curve (see Fig. 6.15) shows that the approach keeps the glucose concentration in the euglycemic range in the preprandial phase (first six hours) like the other approaches. The first meal causes the glucose concentration to rise to a peak of about 300 mg/dL, and it drops only very slightly postprandially. The second meal causes the glucose concentration to rise slightly. The third meal causes the glucose concentration to rise to an average of 500 mg/dL. Postprandially, the glucose concentration drops back to about 300 mg/dL. The last meal then causes the glucose concentration to rise again to an average of 400 mg/dL.

The CVGA (see Fig. 6.16) shows that none of the simulated patients is in the range of risk zones, which indicates a very high glucose variability with a severe risk for all simulated patients.

The simulation results (see Tab. 6.8) show that the algorithm does not respond adequately to the meals taken. Also, looking at the glucose history of the patients (see Fig. 6.15), it can be seen that insulin delivery is not adequately adjusted. The results suggest that the neural network used is not large enough to handle the more than doubled number of possible actions.

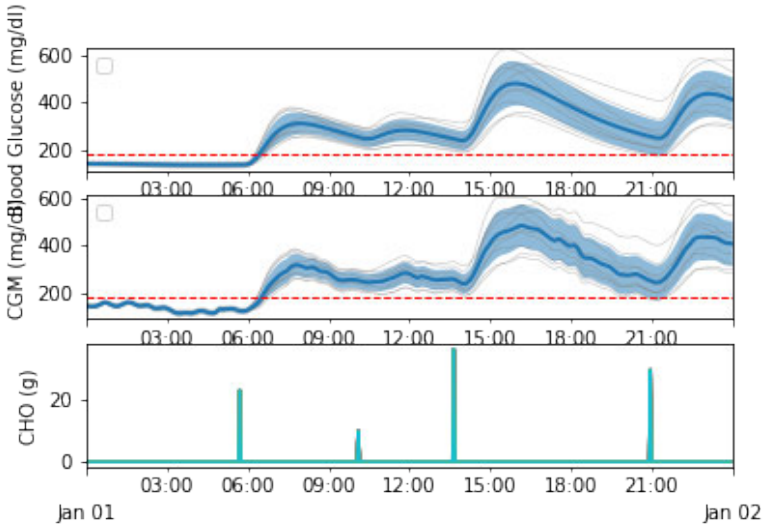

Figure 6.15: Simulation plot (mean values over adult population) of refined action space experiment.

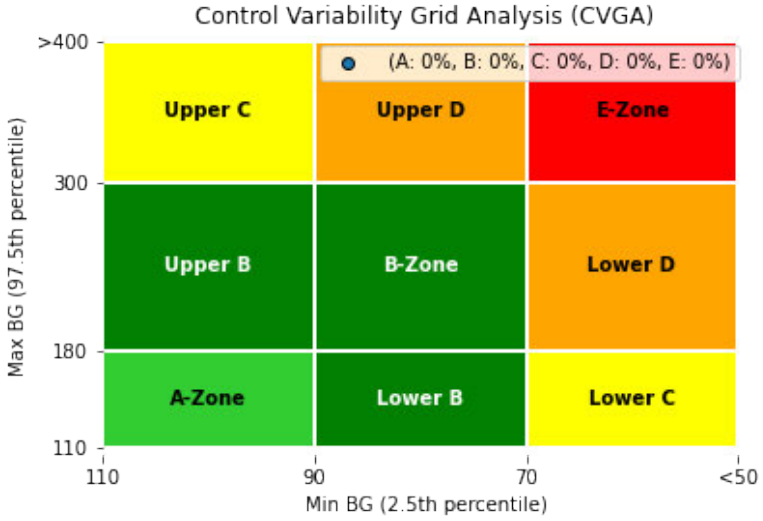

Figure 6.16: CVGA plot of refined action space experiment.

Therefore, in a follow-up experiment, the size of the neural network used should be increased. In order to check whether the enlargement of the network serves as an effect factor at all, we increase the size in a large step to 256 units x 3 layer.

6.5.1 Increase neural network size to 256 units x 3 layer

To investigate the assumption that the increase in the size of the neural network has an effect on the results of the refinement of the action space, in this experiment the neural network size is increased to 256 units x 3 layer. The remaining parameters are unchanged and selected as in the experiment *Using Refined Action Space*.

As the simulation results (see Tab. 6.9) show, the mean Time-In-Range (TIR) across all simulated adult patients is $52.02\% \pm 20.23$, while the mean Time-Above-Range (TAR) is $11.93\% \pm 14.75$ and the mean Time-Below-Range (TBR) is $36.05\% \pm 31.20$.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	71.10	9.56	19.33	5.93	0.89	6.82
adult#002	76.92	5.82	17.26	4.02	0.65	4.67
adult#003	24.74	0.0	75.26	60.43	0.24	60.67
adult#004	37.01	6.24	56.76	25.32	0.09	25.41
adult#005	71.31	28.27	0.42	0.05	5.77	5.82
adult#006	38.88	0.0	61.12	53.84	0.01	53.85
adult#007	21.83	0.0	78.17	120.14	0.0	120.14
adult#008	51.56	0.0	48.44	38.29	0.25	38.54
adult#009	60.71	36.59	2.70	0.0	8.93	8.93
adult#010	66.11	32.85	1.04	0.08	9.03	9.12

Table 6.9: Performance statistics of refined action space experiment (256x3) for adult population.

The simulation course (see Fig. 6.17) shows that the preprandial phase (first six hours) tends slightly toward hypoglycemia. The peaks of glucose concentration that follow the ingested meals are all very close around the upper limit of the euglycemic range. The postprandial phases show that the average glucose concentration is within the euglycemic range. The standard deviation is strikingly large, especially after the third ingested meal.

The CVGA (see Fig. 6.18) shows that 40% of the simulated patients are in one of the D zones, while another 10% are in the Upper-C zone and another 30% are in the Lower-C zone. Another 10% of the patients are in the B zone, while another 10% are in the E zone. This is an improvement to the previously evaluated configuration of 64 units x 3 layer.

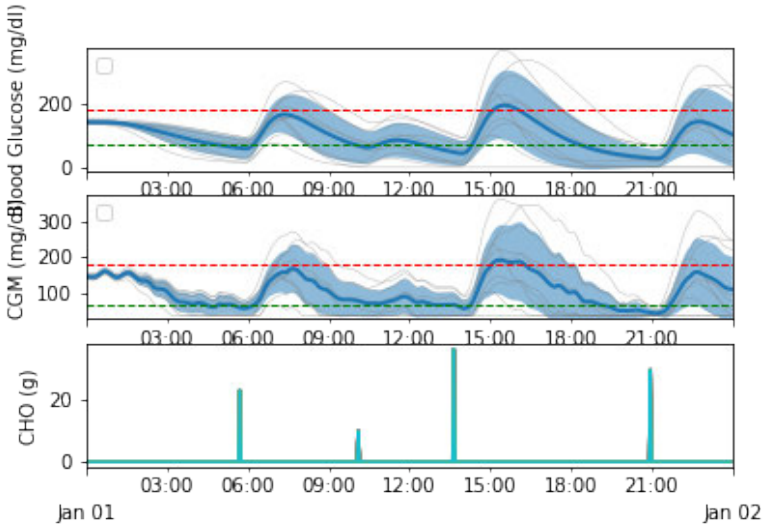

Figure 6.17: Simulation plot (mean values over adult population) of refined action space (256x3) experiment.

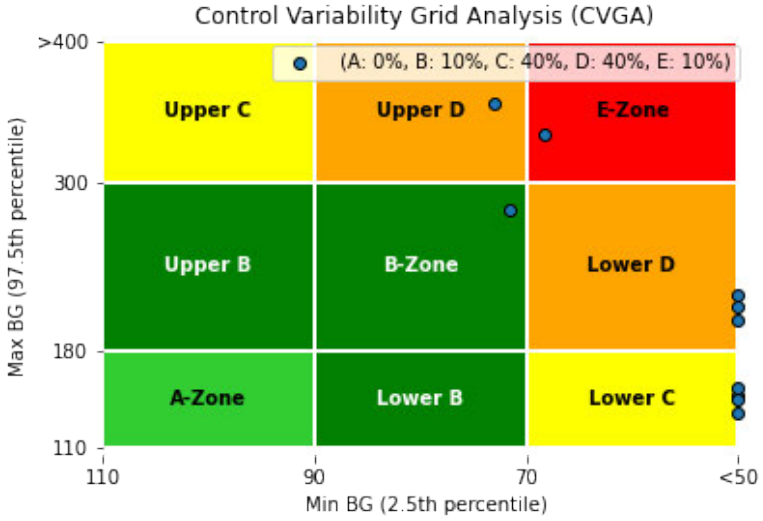

Figure 6.18: CVGA plot of refined action space (256x3) experiment.

Compared to the previous experiment with 64 units x 3 layers, the simulation showed an improvement of the average TIR by 24.93%. The standard deviation has grown by 19.17 compared to the previous experiment.

The high standard deviation suggests that the network tends to overfit due to its size and thus has memorized individual patients.

Possibly, a stronger generalization could be achieved if the neural network is reduced in size again. Therefore, a follow-up experiment with a neural network size of 128 units x 3 layer will be evaluated.

6.5.2 Decrease neural network size to 128 units x 3 layer

The previous experiment has already shown a significant improvement in TIR, but also showed a relatively high standard deviation. The purpose of this experiment is to investigate whether scaling down to a neural network with 128 units x 3 layers shows an improvement in TIR and especially in standard deviation.

As the simulation results (see Tab. 6.10) show, the mean Time-In-Range (TIR) across all simulated adult patients is $57.57\% \pm 7.18$, while the mean Time-Above-Range (TAR) is $35.6\% \pm 10.76$ and the mean Time-Below-Range (TBR) is $6.8\% \pm 6.3$.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	62.58	35.55	1.87	0.00	6.8	6.8
adult#002	67.78	28.69	3.53	0.14	5.79	5.92
adult#003	55.93	31.60	12.47	2.75	9.62	12.37
adult#004	55.72	32.02	12.27	2.07	10.86	12.94
adult#005	59.88	40.12	0.0	0.0	12.07	12.07
adult#006	56.13	30.98	12.89	1.51	6.56	8.07
adult#007	62.16	22.25	15.59	5.29	3.27	8.56
adult#008	63.41	27.23	9.36	0.77	4.55	5.32
adult#009	45.11	54.89	0.0	0.0	16.64	16.64
adult#010	46.99	53.01	0.0	0.0	17.57	17.57

Table 6.10: Performance statistics of refined action space experiment (128x3) for adult population.

The simulation course (see Fig. 6.19) shows that the preprandial phase (first six hours) is maintained within the euglycemic range on average with a very low standard deviation. The first meal is followed by a glucose concentration peak of about 275 mg/dL, while the second meal maintains the glucose concentration within the euglycemic range. The third meal is followed by a glucose peak of about 300-350 mg/dL, but this time the simulated patients do not experience as much postprandial hypoglycemia as in the previous experiments.

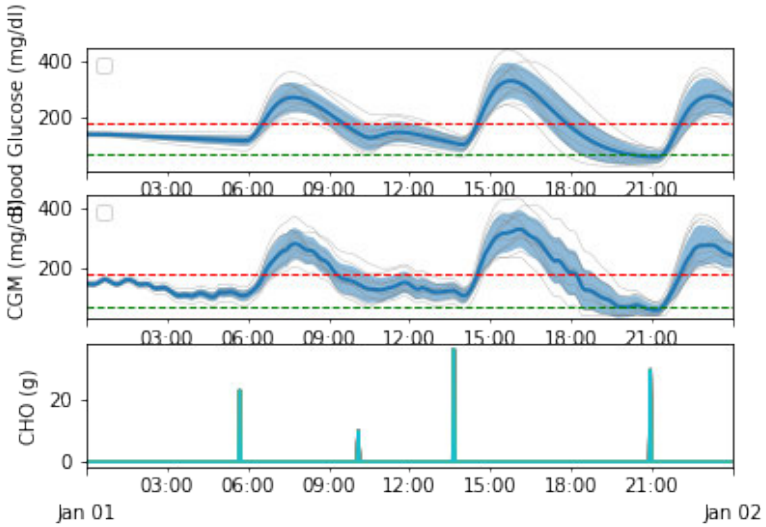

Figure 6.19: Simulation plot (mean values over adult population) of refined action space (128x3) experiment.

The CVGA (see Fig. 6.20) shows that 40% of all patients are in one of the D fields. Another 30% of patients are in the E field. The remaining 30% of patients are in the Upper-C field.

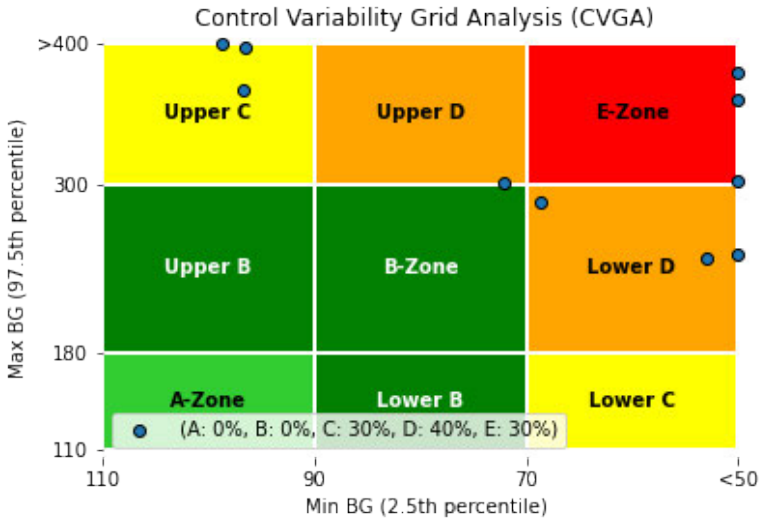

Figure 6.20: CVGA plot of refined action space (128x3) experiment.

Compared to the *IOB (exponential decay, DIA 3h)* experiment, the TIR decreased by an insignificant 0.14%, while the standard deviation decreased by 4.94. The Time-Above-Range (TAR) increased by 12.58%, while the standard deviation decreased by 3.39. The Time-Below-Range (TBR) has decreased by 12.47%, while the standard deviation has decreased by 11.78. It can also be seen in Fig. A.18 that the time with severe hypoglycemia (BG < 50) was significantly reduced. However, it is also observed that severe hyperglycemia cannot be adequately prevented.

As can be seen in the standard deviation in Tab. 6.10, the assumption that reducing the size of the network achieves a stronger generalization appeared to be correct. The standard deviation decreased in all measured metrics. The assumption that a more differentiated action space can better prevent postprandial hypoglycemia has also been confirmed by this experiment.

It is an interesting observation that the TAR increased to such an extent. A follow-up experiment will be conducted to evaluate whether changing the sequence length improves the TIR.

6.6 Different sequence length

So far, the algorithm has learned with batches of size 64 from contiguous sequences of length 32. This corresponds to a sequence length of 96min. The parameter was defined in this way because it was assumed that the long duration of action of insulin can be well represented by this. The following experiments will investigate whether an improvement in the average TIR can be achieved by changing this parameter. This will include looking at how the TAR and TBR compare. The result of the last experiment performed (see Fig. 6.10) serves as a comparison. The remaining parameters were adapted from this experiment.

6.6.1 Decrease sequence length to 16

In this experiment, a first reduction of the sequence length is to be evaluated. For this purpose, the length is set to 16.

As the simulation results (see Tab. 6.11) show, the mean Time-In-Range (TIR) across all simulated adult patients is $51.44\% \pm 14.98$, while the mean Time-Above-Range (TAR) is $11.02\% \pm 13.77$ and the mean Time-Below-Range (TBR) is $37.55\% \pm 25.27$.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	63.83	5.2	30.98	20.67	0.43	21.1
adult#002	67.15	0.0	32.85	14.22	0.12	14.34
adult#003	41.37	6.86	51.77	66.35	0.39	66.74
adult#004	43.45	8.32	48.23	17.75	0.37	18.12
adult#005	69.02	22.87	8.11	0.48	3.48	3.96
adult#006	42.62	0.0	57.38	57.32	0.06	57.38
adult#007	25.99	0.0	74.01	110.24	0.01	110.26
adult#008	37.42	0.0	62.58	46.5	0.38	46.88
adult#009	59.46	36.17	4.37	0.0	9.03	9.03
adult#010	64.03	30.77	5.2	0.33	7.15	7.47

Table 6.11: Performance statistics of different sequence length (16) for adult population.

The simulation course (see Fig. 6.21) shows that the ingested meal produces a lower peak in glucose concentration than in the configuration with a sequence length of 32. However, in the postprandial phases, the simulated patients tend to experience hypoglycemia. This is also reflected in the high TBR.

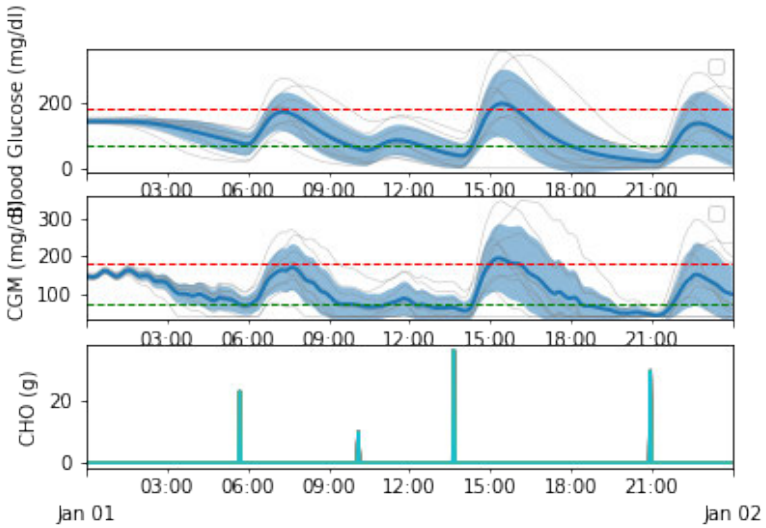

Figure 6.21: Simulation plot (mean values over adult population) of different sequence length (16) experiment.

The CVGA (see Fig. 6.22) shows that 40% of patients are in one of the C fields, while another 40% are in one of the D fields. Another 20% are in the E field.

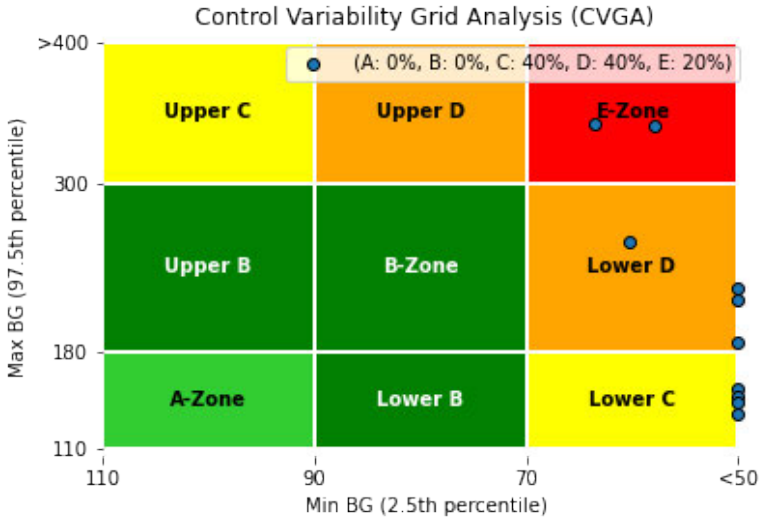

Figure 6.22: CVGA plot of different sequence length (16) experiment.

Compared with the configuration with a sequence length of 32, more patients (+10%) are now located in one of the less risky C fields, but patients are located at the outer edge of the C and D fields, which correlates with the observed hypoglycemias.

The observed hypoglycemias should be critically evaluated, as the algorithm cannot provide the necessary safety for clinical use.

In the further course, a stronger reduction of the sequence length will be evaluated. The aim is to check whether the observed hypoglycemias can be prevented and how the average TIR behaves as a result of this change.

6.6.2 Decrease sequence length to 8

After the previous experiment demonstrated that the reduction of the sequence length does not lead to any improvement of the average TIR, this experiment will evaluate a further reduction of the sequence length to length 8.

As the simulation results (see Tab. 6.12) show, the mean Time-In-Range (TIR) across all simulated adult patients is $56.24\% \pm 14.98$, while the mean Time-Above-Range (TAR) is $29.75\% \pm 18.01$ and the mean Time-Below-Range (TBR) is $14.01\% \pm 11.82$.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	66.53	18.71	14.76	1.65	2.25	3.89
adult#002	78.17	11.23	10.60	1.28	1.73	3.01
adult#003	58.42	24.74	16.84	0.96	3.79	4.75
adult#004	54.05	28.07	17.88	0.95	5.12	6.08
adult#005	67.78	32.22	0.0	0.0	8.06	8.06
adult#006	47.40	29.11	23.49	5.4	5.97	11.37
adult#007	37.21	26.2	36.59	10.03	4.85	14.88
adult#008	73.39	6.65	19.96	3.98	1.35	5.32
adult#009	37.63	62.37	0.0	0.0	21.26	21.26
adult#010	41.79	58.21	0.0	0.0	20.75	20.75

Table 6.12: Performance statistics of different sequence length experiment (8) for adult population.

The simulation course (see Fig. 6.23) shows that the ingested meals lead to stronger peaks in glucose concentration than in the configuration with a sequence length of 16. It can be seen that the postprandial phases are on average within the euglycemic range. However, some simulated patients also experience more severe hypoglycemia.

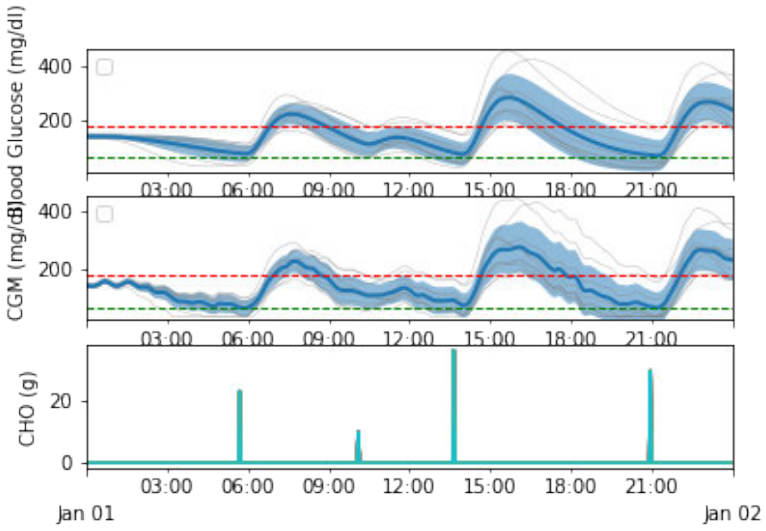

Figure 6.23: Simulation plot (mean values over adult population) of different sequence length (8) experiment.

The CVGA (see Fig. 6.24) shows that 60% of the simulated patients are in one of the D fields, while another 10% are in the Upper B field. Another 10% are in the E field.

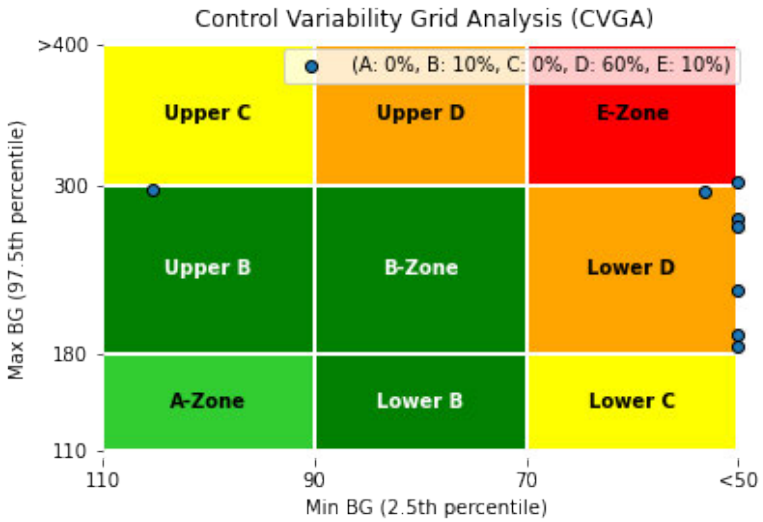

Figure 6.24: CVGA plot of different sequence length (8) experiment.

Compared to the previous experiment, 20% fewer patients lie within the range of the CVGA. Also in this experiment, patients often lie at the right edge of the CVGA, which correlates with the observed hypoglycemias.

Since both reductions of the sequence length did not result in significant improvements of the measured metrics, an increase of the sequence length to 64 will be evaluated in a follow-up experiment.

6.6.3 Increase sequence length to 64

In this experiment, an increase of the sequence length to a length of 64 will be evaluated. It will be observed whether the postprandial hypoglycemias that occurred previously can be avoided.

As the simulation results (see Tab. 6.13) show, the mean Time-In-Range (TIR) across all simulated adult patients is $53.31\% \pm 11.59$, while the mean Time-Above-Range (TAR) is $42.37\% \pm 12.26$ and the mean Time-Below-Range (TBR) is $4.32\% \pm 5.98$.

The simulation course (see Fig. 6.25) shows that the ingested meals lead to relatively strong peaks of glucose concentration. The postprandial phases show good glucose control in the euglycemic range.

Patient	TIR (%)	TAR (%)	TBR (%)	LBGI	HBGI	BGRI
adult#001	56.96	43.04	0.0	0.0	8.77	8.77
adult#002	66.74	33.26	0.0	0.0	6.42	6.42
adult#003	49.27	40.12	10.60	2.02	12.09	14.12
adult#004	50.52	37.84	11.64	2.16	13.1	15.25
adult#005	42.0	58.00	0.0	0.0	15.36	15.36
adult#006	62.16	31.81	6.03	0.05	8.16	8.21
adult#007	48.86	36.17	14.97	2.62	5.99	8.61
adult#008	72.97	27.03	0.0	0.0	5.76	5.76
adult#009	34.30	65.7	0.0	0.0	22.72	22.72
adult#010	49.27	50.73	0.0	0.0	15.55	15.55

Table 6.13: Performance statistics of different sequence length experiment (64) for adult population.

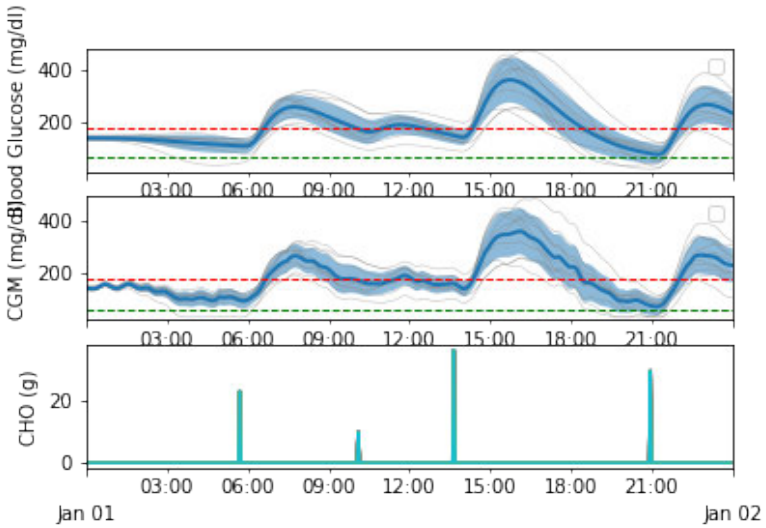

Figure 6.25: Simulation plot (mean values over adult population) of different sequence length (64) experiment.

The CVGA (see Fig. 6.26) shows that 30% of patients are located in the Upper-C field, while 10% of patients are located in the Lower-D field. Another 30% are located in the E field. Another 10% are located in the Upper-B field.

Compared with the previous experiment, hypoglycemia was successfully avoided in this configuration (with a TBR of 4.32%±5.98). However, the risk assessment also shows that the risk spread of the patients is relatively high. The TIR dropped by 4.99% compared

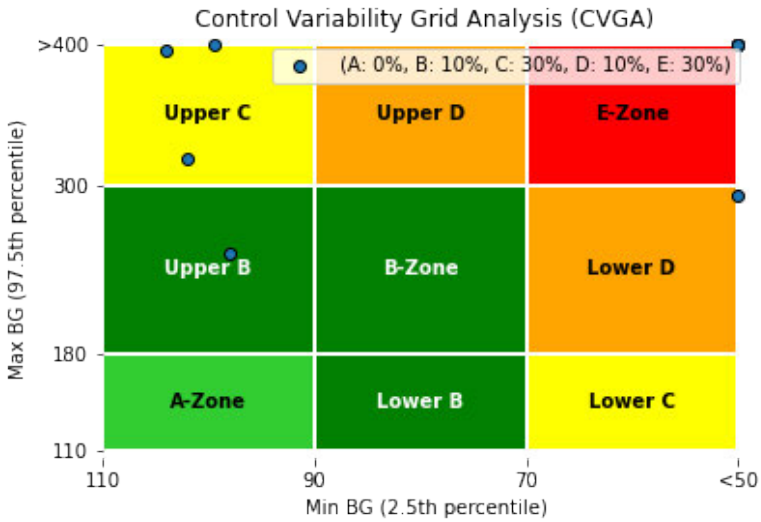

Figure 6.26: CVGA plot of different sequence length (64) experiment.

to the configuration with a sequence length of 32, while the standard deviation shows a slight improvement of -1.24.

However, this approach appears to provide no significant improvement compared to the first sequence length configuration of 32.

6.7 Results

In this section, the results of all experiments performed are summarized in tabular form (see Tab. 6.14). For this purpose, the mean TIR, TAR and TBR achieved are presented with their respective standard deviation.

The result determined for comparison in the upcoming discussion is the evaluation result *Refined action space (128x3)*, since a high TIR and a low TBR with a low standard deviation were measured here.

Experiment	Mean TIR (%)	Mean TAR (%)	Mean TBR (%)
PID controller	46.84 ± 13.29	15.8 ± 5.50	37.36 ± 14.15
Baseline implementation	52.24 ± 15.81	46.22 ± 18.37	1.54 ± 4.86
Applied scaling	50.62 ± 12.90	15.55 ± 11.32	33.83 ± 19.76
IOB (bilinear)	48.02 ± 27.95	48.77 ± 31.79	3.20 ± 6.98
IOB (exponential, DIA 3h)	57.71 ± 12.12	23.02 ± 14.15	19.27 ± 18.08
IOB (exponential, DIA 5h)	40.35 ± 13.04	53.62 ± 23.33	6.03 ± 16.99
Refined action space (64x3)	27.09 ± 1.06	72.91 ± 1.06	0.00 ± 0.00
Refined action space (256x3)	52.02 ± 20.23	11.93 ± 14.75	36.05 ± 31.20
Refined action space (128x3)	57.57 ± 7.18	35.6 ± 10.76	6.8 ± 6.3
Different sequence length (16)	51.44 ± 14.98	11.02 ± 13.77	37.55 ± 25.27
Different sequence length (8)	56.24 ± 14.98	29.75 ± 18.01	14.01 ± 11.82
Different sequence length (64)	53.31 ± 11.59	42.37 ± 12.26	4.32 ± 5.98

Table 6.14: Overview of all experiment results.

7 Discussion

In this chapter, the previously obtained evaluation results will be used to answer the research questions stated at the beginning of this thesis. Furthermore, the results will be put into the context of the research area and possible differences and weaknesses will be discussed.

The goal of this thesis was to answer how a Deep Q-Learning approach needs to be designed to meet the requirements of good diabetes therapy solely through basal rate adjustment despite the different metabolic factors of individual patients. A Deep Q-Learning approach was presented, with evaluation results showing that the achieved TIR of 57.57%±7.18 is within the target range of good diabetes therapy. All experiments presented were performed with basal rate adjustment only. No meal boluses were performed or sports sessions included. Nevertheless, it is suggested that additional meal boluses could significantly improve the achieved TIR and thus allow for better diabetes therapy. The low standard deviation of 7.18 indicates that a generalization of the algorithm has been achieved that can handle the differences in the metabolic rates of individual patients. Thus, a basically good generalization can be observed in the results, but the measured TIR is below 50% in some simulated patients (45.11% in the worst case).

Also the question to what extent a good diabetes therapy can be ensured despite deviations from regular eating patterns was to be answered in the scope of this thesis. Since the presented approach was trained and simulated with a meal scenario randomly distributed around a regular pattern, the proposed algorithm is capable of dealing with these situations within the definition of a good diabetes therapy. Nevertheless, it should be noted that this statement can only be made to a limited extent, since the experiments were only simulated over a period of 24 hours. Thus, the ingested meals occurred only once initially by chance and not over several days. However, the simulated 24h still contain randomly scattered meals that are reliably answered by the algorithm. It should also be noted that no conclusion can be made about how much the deviations from the meal scenario may be scattered, so that the algorithm still meets the goals of good diabetes

therapy. Further research is needed in this regard. Another interesting aspect that has been read about in other research papers is the omission of a meal. In the context of this thesis, no experiments were conducted that could answer how the algorithm would handle a skipped meal. Further research is needed for this as well.

Another research question asked is how the developed approach can act safely and avoid severe hypoglycemia. The evaluation results of the presented approach indicate that severe hypoglycemia is omitted. The obtained TBR of $6.8\% \pm 6.3$ is in a very low range. The standard deviation is also very low, which supports this assumption. Nevertheless, severe hypoglycemia cannot be completely avoided in individual simulated patients. It would be conceivable to avoid these even better by an emergency deactivation of the basal rate in the event of an excessive drop in glucose concentration. However, it has been noted that when glucose peaks have been avoided, the risk of hypoglycemia has increased significantly. In some experiments the algorithm adjusted the insulin delivery to aim for preprandial hypoglycemias. This led to slightly less high peaks of glucose concentration after meal intake, but indicates that higher glucose peaks needs to be tolerated to avoid severe hypoglycemia. An adjustment of the reward function to stronger penalize hypoglycemias could lead to an approvement of this behaviour. Further research needs to be conducted in this regard.

It was also to be evaluated what the impact of using Insulin-On-Board (IOB) as a factor of the algorithm to be developed is. Insulin On Board (IOB) as a factor has been shown to make a significant difference in the success of the algorithm applied. The difference in measured TIR (%) from 50.62 ± 12.9 without IOB to 57.71 ± 12.12 with IOB represents a substantial improvement in the control strategy. In particular, the occurrence of the severely dangerous hypoglycemias could be significantly reduced by the inclusion of IOB as a factor. This can be observed by the difference in TBR (%) of 33.83 ± 19.76 without IOB to 19.27 ± 18.08 with IOB. However, it is important to ensure that the appropriate modeling of pharmacodynamics is used. Bilinear modeling has been shown to worsen TIR, while exponential modeling significantly improves results in the control strategy. For correct use of IOB modeling, the DIA and peak must be known as parameters, each of which depends on the type of insulin used. Thus, if the insulin type and its characteristics are not known in advance, omitting Insulin-On-Board as a factor may allow for a better control strategy.

Another question addressed was how effectively the algorithm achieves the goals of good diabetes therapy even with the consumption of high-carb meals. The meal scenario used

contains meals in the range of 30g to 110g carbohydrates, where 110g is considered as a high-carb meal. The goals of good diabetes therapy are met by the measured TIR of 57.57%±7.18. However, it has been observed that when glucose peaks have been avoided, the risk of hypoglycemia has increased noticeably. To avoid severe hypoglycemia, higher glucose peaks must be tolerated as previously stated. The glucose concentration peaks, however, might be better dealt with the use of meal boluses, and thus result in better TIR and therefore better diabetes therapy could be achieved. This can also be observed in the non-optimal risk scores in the respective CVGA plots. Only a few experiments scored in the less risky category A or B fields, while many were at best in categories D. This is mainly due to glucose variability that presumably could be significantly reduced by preprandial meal boluses.

Unfortunately, comparison with other approaches appears to be very difficult due to the often very different metrics used and experiment setups selected. For example, the approach by Sun et al. [90] uses a much less challenging meal scenario of 07hr (50g), 12hr (60g), 18:30 (80g), 23hr (16g) with variability of only ±15 minutes. In addition, this approach does not deliver bolus insulin only for the last meal (11pm, 16g), which means that the approach is more likely to be considered as hybrid loop. It is likely that the approach presented in this thesis would yield better results under the simulation conditions of Sun et al. This would be interesting to investigate in a follow-up work. The approach by Lee et al. [53] also uses a much less challenging meal scenario of 08am (40g), 1pm (80g) and 9pm (60g). In fact, one less meal was simulated here. Above all, the lower variability of the related approaches may be partly responsible for the significantly better results. Additionally, no variability in meal times was used. The approach by Sanz et al. [82] also uses a much less challenging meal scenario of 07hr (65g), 14hr (70g), 21hr (65g). They use a standard deviation of 60min, analogous to the presented approach in this thesis. However, the variability in the meals is significantly lower in comparison. The approach by Rossetti et al. [81] was studied exclusively in a clinical trial investigating the use of only one 60g meal. A bolus was delivered for the meal, which characterizes the approach as hybrid loop.

Another parameter that is often chosen very diversely is the length of the simulation period in the evaluation. Sun et al. [90] simulated a period of 98 days, while Sanz et al. [82] used a period of 30 days and Lee et al. [53], for example, only simulated a period of 24h. This further does not contribute to the comparability of the results. The approach presented in this thesis was also evaluated with a time period of 24h. With a longer simulation period, stronger statements could be made about the stability and potential

Approach	TIR (%)	TAR (%)	TBR (%)
Sun et al. [90]	89.9 ± 8.7	6.1 ± 8.2	2.5 ± 3.0
Lee et al. [53]	89.56 ± 4.37	9.49 ± 4.28	0.95 ± 1.97
Sanz et al. [82]	80.0	1.3*	n/a
Rossetti et al. [81]	78.8	n/a	5.2
Atlas et al. [4]	73.0	27.0	0.0
Presented approach	57.57 ± 7.18	35.6 ± 10.76	6.8 ± 6.3
PID Controller	40.02 ± 8.11	20.1 ± 4.05	39.88 ± 9.73

Table 7.1: Comparison of approaches (* TAR defined as BG $>$ 250).

applicability of the algorithm. However, this was not feasible within the scope of this thesis due to time constraints.

Another difference is the choice of metrics for presenting the results. In the paper of Sanz et al. [82] only a TIR and a TAR are given. The specification for the TBR is missing. In addition, the TAR was defined differently than in other papers. It is specified as BG $>$ 250 mg/dL here. In the work by Rossetti et al. [81] the data on the TAR are missing. Only the TIR and the TBR are given.

In Tab. 7.1 the simulation results are presented in comparison to the approaches presented in the third chapter and put into the context of the related work. The final result of the proposed algorithm achieves a TIR (%) of 57.57 ± 7.18, which indicates a better control strategy and diabetes therapy than the presented PID controller approach. The other approaches presented achieve higher TIR results. Among other reasons, this is due to the difficult comparability caused by the varying experimental setups as already stated.

Unfortunately, it has not been possible to apply the algorithm to the data already available from the OpenAPS Foundation. The reasons for this are on the one hand that the data preparation would have been too time-consuming for the scope of this work, as this would have required completely clean data sets with a continuous glucose history and, above all, data sets with consistently announced meals. Specifically, this would mean that the replay memory would be filled with the collected glucose histories, including meal information and associated insulin deliveries, which would then serve as the replay memory for the algorithm. An implementation with the real patient data could provide interesting insights into whether the algorithm already trained on the simulated data will also perform well on the real data.

The presented algorithm has been trained using only simulated data from the patient *adult*#001. It would be worth investigating whether the results in TIR would improve if the training was also performed on the other simulated patients.

Also, varying influencing factors such as sports or stress are currently not simulated by the simulator, which means that the developed algorithm was not confronted with such situations and could not learn from them. First hints how such a consideration could take place and which problems can occur in this context can be found in a paper by Tagougui et al. [93].

In the approach by Sanz et al. [82], among other features, a safety mechanism is implemented that switches off the insulin supply under certain conditions. Since the algorithm may not have learned certain situations because they did not occur in the learning process, safety measures in the form of insulin cutoff at critically low glucose concentration similar to this approach could be implemented if the algorithm is prepared for use in subsequent clinical trials.

The results are a first step towards the development of an artificial pancreas. However, since the experiments were conducted exclusively in-silico, no statement can yet be made as to whether the system can actually achieve significant therapy improvements in real patients. The individual metabolic processes can only be simulated by the simulator and thus can only ever be understood as an approximation to real conditions. In order to be able to make statements about the actual relevance of the approach, in-vivo studies are necessary.

8 Outlook

In this thesis an algorithm is presented that meets the requirements for good diabetes therapy solely by adjusting the basal rate. It can also deal with the different metabolic factors of individual patients. It has been shown that deviations from regular eating patterns can be responded to with this algorithm. However, it was also found that the glucose peaks cannot be prevented if the focus is on avoiding severe hypoglycemia. The presented algorithm is able to act safely and avoid severe hypoglycemia, if postprandial glucose peaks are tolerated. However, it has been found that the algorithm is not capable of avoiding severe hypoglycemia in all simulated patients. For this purpose the need of a further IOB limitation is suspected to support this goal. The use of Insulin-On-Board (IOB) as an input factor for the algorithm has been shown to be helpful if the correct modeling is known and specified by the type of insulin used. If this is not the case, the therapy goals are achieved less well than without the inclusion of IOB. When high-carb meals are taken, it has been found that glucose peaks are not preventable if postprandial hypoglycemia is to be avoided.

Following from the results of this thesis it is planned to apply the implementation of the algorithm, as described beforehand, to real patient data. The use of the provided patient data from the *OpenAPS Data Commons* program is linked to the obligation to share the achieved research results with the community. For this purpose, a paper summarizing the results will be written and published.

Since an implementation of the approach presented here in an Android application exceeded the scope of this work, it is planned to implement the algorithm directly as a preparation for planned clinical trials.

Within the scope of this work, a rudimentary hyperparameter tuning has been carried out, which is to be optimized under further expenditure of time. Hereby it might be possible to further improve the already achieved results.

During the conduct of the research, opportunities for improvement in the UVa/Padova implementation Simglucose have repeatedly come to attention, leading to active collaboration on the code base of the simulator. This collaboration will be intensified in the future to make research using the open-source simulator more accessible and thus provide a lower-threshold entry into the development of insulin delivery algorithms.

Unfortunately, circumstances during the execution of the work did not allow to use the latest approved metabolic simulator (UVa/Padova S2013). The improvement of the metabolic models in the newer version could significantly increase the accuracy of the presented algorithm. This will be evaluated following this work.

In order to achieve the long-term goal of an intervention-free diabetes therapy through the development of an artificial pancreas, some further investigations have to be carried out in the future. The approvability of algorithms based on machine learning represents a great challenge, since such algorithms lack the traceability of decisions. It would be interesting to investigate whether there are already existing solutions for the problem of *Explainable AI* in the context of Artificial Pancreas.

It would also be very interesting to combine the algorithm with an unannounced meal detection algorithm already developed in previous work to completely replace the announcement of meals. Thus, a completely intervention-free system would be conceivable.

Furthermore, new CGM systems have come onto the market that have a significantly better error deviation. Also systems with a higher measurement frequency (one minute instead of three minutes) are now available on the market. It would be interesting to investigate whether the presented algorithm can improve therapy under these conditions.

Based on the work presented here, further research is planned as part of a dissertation on the topic of Artificial Pancreas. The dissertation is to be conducted within the framework of the non-medical PhD program of the University Medical Center Hamburg-Eppendorf (UKE). This framework will allow the algorithm presented here to be clinically tested through patient studies. In addition, the challenges described in the previous paragraph will be explored in this framework.

It is also planned to deepen the research in the area of approvability as a medical device of reinforcement learning based approaches in the context of Artificial Pancreas.

So the way to a complete artificial pancreas and thus an intervention-free diabetes therapy still holds some stumbling blocks and interesting questions. This thesis contributes to the development of such a system.

Bibliography

[1] Advani, Andrew: Positioning time in range in diabetes management. In: *Diabetologia* 63 (2020), Nr. 2, S. 242–252

[2] Anderson, Ryan J. ; Freedland, Kenneth E. ; Clouse, Ray E. ; Lustman, Patrick J.: The prevalence of comorbid depression in adults with diabetes: a meta-analysis. In: *Diabetes care* 24 (2001), Nr. 6, S. 1069–1078

[3] Asad, Muhammad ; Qamar, Usman: A review of continuous blood glucose monitoring and prediction of blood glucose level for diabetes type 1 patient in different prediction horizons (ph) using artificial neural network (ann). In: *Proceedings of SAI Intelligent Systems Conference* Springer (Veranst.), 2019, S. 684–695

[4] Atlas, Eran ; Nimri, Revital ; Miller, Shahar ; Grunberg, Eli A. ; Phillip, Moshe: MD-logic artificial pancreas system: a pilot study in adults with type 1 diabetes. In: *Diabetes care* 33 (2010), Nr. 5, S. 1072–1076

[5] Austenat, Elke ; Stahl, Tilman: *Insulin Pump Therapy: Indication-Method-Technology*. Walter de Gruyter GmbH & Co KG, 2019

[6] Bakhtiani, Parkash A. ; Zhao, Lauren M. ; El Youssef, Joseph ; Castle, Jessica R. ; Ward, W K.: A review of artificial pancreas technologies with an emphasis on bi-hormonal therapy. In: *Diabetes, Obesity and Metabolism* 15 (2013), Nr. 12, S. 1065–1070

[7] Bamgbose, Samuel O. ; Li, Xiangfang ; Qian, Lijun: Closed loop control of blood glucose level with neural network predictor for diabetic patients. In: *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)* IEEE (Veranst.), 2017, S. 1–6

[8] Barnard, KD ; Skinner, TC ; Peveler, R: The prevalence of co-morbid depression in adults with Type 1 diabetes: systematic literature review. In: *Diabetic medicine* 23 (2006), Nr. 4, S. 445–448

[9] Battelino, Tadej ; Danne, Thomas ; Bergenstal, Richard M. ; Amiel, Stephanie A. ; Beck, Roy ; Biester, Torben ; Bosi, Emanuele ; Buckingham, Bruce A. ; Cefalu, William T. ; Close, Kelly L. u. a.: Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. In: *Diabetes care* 42 (2019), Nr. 8, S. 1593–1603

[10] Bekiari, Eleni ; Kitsios, Konstantinos ; Thabit, Hood ; Tauschmann, Martin ; Athanasiadou, Eleni ; Karagiannis, Thomas ; Haidich, Anna-Bettina ; Hovorka, Roman ; Tsapas, Apostolos: Artificial pancreas treatment for outpatients with type 1 diabetes: systematic review and meta-analysis. In: *bmj* 361 (2018)

[11] Bequette, B W.: A critical assessment of algorithms and challenges in the development of a closed-loop artificial pancreas. In: *Diabetes technology & therapeutics* 7 (2005), Nr. 1, S. 28–47

[12] Bequette, B W.: Challenges and progress in the development of a closed-loop artificial pancreas. In: *2012 American Control Conference (ACC)* IEEE (Veranst.), 2012, S. 4065–4071

[13] Bequette, B W.: Algorithms for a closed-loop artificial pancreas: the case for model predictive control. In: *Journal of diabetes science and technology* 7 (2013), Nr. 6, S. 1632–1643

[14] Berián, Jesús ; Bravo, Ignacio ; Gardel-Vicente, Alfredo ; Lázaro-Galilea, José-Luis ; Rigla, Mercedes: Dynamic Insulin Basal Needs Estimation and Parameters Adjustment in Type 1 Diabetes. In: *Sensors* 21 (2021), Nr. 15, S. 5226

[15] Biewald, Lukas: *Experiment Tracking with Weights and Biases*. 2020. – URL https://www.wandb.com/. – Software available from wandb.com

[16] Blauw, H ; Van Bon, AC ; Koops, R ; DeVries, JH ; consortium, PCDIAB: Performance and safety of an integrated bihormonal artificial pancreas for fully automated glucose control at home. In: *Diabetes, Obesity and Metabolism* 18 (2016), Nr. 7, S. 671–677

[17] Bock, Martin I. de ; Roy, Anirban ; Cooper, Matthew N. ; Dart, Julie A. ; Berthold, Carolyn L. ; Retterath, Adam J. ; Freeman, Kate E. ; Grosman, Benyamin ; Kurtz, Natalie ; Kaufman, Fran u. a.: Feasibility of outpatient 24-hour closed-loop insulin delivery. In: *Diabetes Care* 38 (2015), Nr. 11, S. e186–e187

[18] Bothe, Melanie K. ; Dickens, Luke ; Reichel, Katrin ; Tellmann, Arn ; Ellger, Björn ; Westphal, Martin ; Faisal, Ahmed A.: The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. In: *Expert review of medical devices* 10 (2013), Nr. 5, S. 661–673

[19] Boughton, Charlotte K. ; Hovorka, Roman: The artificial pancreas. In: *Current Opinion in Organ Transplantation* 25 (2020), Nr. 4, S. 336–342

[20] Bradley, Clare ; Speight, Jane: Patient perceptions of diabetes and diabetes therapy: assessing quality of life. In: *Diabetes/Metabolism Research and Reviews* 18 (2002), Nr. S3, S. S64–S69

[21] Brewer, Karl W. ; Chase, H P. ; Owen, Susie ; Garg, Satish K.: Slicing the pie: correlating HbA1c values with average blood glucose values in a pie chart form. In: *Diabetes Care* 21 (1998), Nr. 2, S. 209–212

[22] Broome, David T. ; Hilton, C B. ; Mehta, Neil: Policy implications of artificial intelligence and machine learning in diabetes management. In: *Current diabetes reports* 20 (2020), Nr. 2, S. 1–5

[23] Brown, Sue A. ; Kovatchev, Boris P. ; Breton, Marc D. ; Anderson, Stacey M. ; Keith-Hynes, Patrick ; Patek, Stephen D. ; Jiang, Boyi ; Ben Brahim, Najib ; Vereshchetin, Paul ; Bruttomesso, Daniela u. a.: Multinight "bedside" closed-loop control for patients with type 1 diabetes. In: *Diabetes technology & therapeutics* 17 (2015), Nr. 3, S. 203–209

[24] Cinar, Ali ; Turksoy, Kamuran: *Advances in Artificial Pancreas Systems: Adaptive and Multivariable Predictive Control*. Springer, 2018

[25] Cobelli, Claudio ; Renard, Eric ; Kovatchev, Boris: Artificial pancreas: past, present, future. In: *Diabetes* 60 (2011), Nr. 11, S. 2672–2682

[26] Control, Diabetes ; Group, Complications Trial R.: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. In: *New England journal of medicine* 329 (1993), Nr. 14, S. 977–986

[27] Da Silva, Julien ; Lepore, Giuseppe ; Battelino, Tadej ; Arrieta, Arcelia ; Castañeda, Javier ; Grosman, Benyamin ; Shin, John ; Cohen, Ohad: Real-world Performance of the MiniMed™ 780G System: First Report of Outcomes from 4'120 Users. In: *Diabetes Technology and Therapeutics* (2021), Nr. ja

[28] Dai, Xia ; Luo, Zu-chun ; Zhai, Lu ; Zhao, Wen-piao ; Huang, Feng: Artificial pancreas as an effective and safe alternative in patients with type 1 diabetes mellitus: a systematic review and meta-analysis. In: *Diabetes Therapy* 9 (2018), Nr. 3, S. 1269–1277

[29] Danne, Thomas ; Nimri, Revital ; Battelino, Tadej ; Bergenstal, Richard M. ; Close, Kelly L. ; DeVries, J H. ; Garg, Satish ; Heinemann, Lutz ; Hirsch, Irl ; Amiel, Stephanie A. u. a.: International consensus on use of continuous glucose monitoring. In: *Diabetes care* 40 (2017), Nr. 12, S. 1631–1640

[30] De Bois, Maxime ; El Yacoubi, Mounîm A ; Ammi, Mehdi: Prediction-coherent LSTM-based recurrent neural network for safer glucose predictions in diabetic people. In: *International Conference on Neural Information Processing* Springer (Veranst.), 2019, S. 510–521

[31] Doyle, Francis J. ; Huyett, Lauren M. ; Lee, Joon B. ; Zisser, Howard C. ; Dassau, Eyal: Closed-loop artificial pancreas systems: engineering the algorithms. In: *Diabetes care* 37 (2014), Nr. 5, S. 1191–1197

[32] Ehrmann, Dominic ; Kulzer, Bernhard ; Haak, Thomas ; Hermanns, Norbert: Longitudinal relationship of diabetes-related distress and depressive symptoms: analysing incidence and persistence. In: *Diabetic Medicine* 32 (2015), Nr. 10, S. 1264–1271

[33] El-Khatib, Firas H. ; Russell, Steven J. ; Nathan, David M. ; Sutherlin, Robert G. ; Damiano, Edward R.: A bihormonal closed-loop artificial pancreas for type 1 diabetes. In: *Science translational medicine* 2 (2010), Nr. 27, S. 27ra27–27ra27

[34] Farrington, Conor: Hacking diabetes: DIY artificial pancreas systems. In: *The Lancet Diabetes & Endocrinology* 5 (2017), Nr. 5, S. 332

[35] Forlenza, Gregory P. ; Li, Zoey ; Buckingham, Bruce A. ; Pinsker, Jordan E. ; Cengiz, Eda ; Wadwa, R P. ; Ekhlaspour, Laya ; Church, Mei M. ; Weinzimer, Stuart A. ; Jost, Emily u. a.: Predictive low-glucose suspend reduces hypoglycemia in adults, adolescents, and children with type 1 diabetes in an at-home randomized crossover study: results of the PROLOG trial. In: *Diabetes care* 41 (2018), Nr. 10, S. 2155–2161

[36] Fox, Ian ; Lee, Joyce ; Pop-Busui, Rodica ; Wiens, Jenna: Deep reinforcement learning for closed-loop blood glucose control. In: *Machine Learning for Healthcare Conference* PMLR (Veranst.), 2020, S. 508–536

[37] Fox, Ian ; Wiens, Jenna: Reinforcement learning for blood glucose control: Challenges and opportunities. (2019)

[38] Group, Juvenile Diabetes Research Foundation Continuous Glucose Monitoring S.: Continuous glucose monitoring and intensive treatment of type 1 diabetes. In: *New England Journal of Medicine* 359 (2008), Nr. 14, S. 1464–1476

[39] Haidar, Ahmad: Insulin-and-glucagon artificial pancreas versus insulin-alone artificial pancreas: A short review. In: *Diabetes Spectrum* 32 (2019), Nr. 3, S. 215–221

[40] Han, Jianchao ; Rodriguez, Juan C. ; Beheshti, Mohsen: Diabetes data analysis and prediction model discovery using rapidminer. In: *2008 Second international conference on future generation communication and networking* Bd. 3 IEEE (Veranst.), 2008, S. 96–99

[41] Hochreiter, Sepp ; Schmidhuber, Jürgen: Long short-term memory. In: *Neural computation* 9 (1997), Nr. 8, S. 1735–1780

[42] Hoover, Joan W.: Patient burnout, and other reasons for noncompliance. In: *The Diabetes Educator* 9 (1983), Nr. 3, S. 41–43

[43] Hovorka, Roman ; Canonico, Valentina ; Chassin, Ludovic J. ; Haueter, Ulrich ; Massi-Benedetti, Massimo ; Federici, Marco O. ; Pieber, Thomas R. ; Schaller, Helga C. ; Schaupp, Lukas ; Vering, Thomas u. a.: Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. In: *Physiological measurement* 25 (2004), Nr. 4, S. 905

[44] Jiang, Fei ; Jiang, Yong ; Zhi, Hui ; Dong, Yi ; Li, Hao ; Ma, Sufeng ; Wang, Yilong ; Dong, Qiang ; Shen, Haipeng ; Wang, Yongjun: Artificial intelligence in healthcare: past, present and future. In: *Stroke and vascular neurology* 2 (2017), Nr. 4

[45] Karageorgiou, Vasilios ; Papaioannou, Theodoros G. ; Bellos, Ioannis ; Alexandraki, Krystallenia ; Tentolouris, Nikolaos ; Stefanadis, Christodoulos ; Chrousos, George P. ; Tousoulis, Dimitrios: Effectiveness of artificial pancreas in the non-adult population: a systematic review and network meta-analysis. In: *Metabolism* 90 (2019), S. 20–30

[46] Kerner, Wolfgang ; Brückel, J: Definition, classification and diagnosis of diabetes mellitus. In: *Experimental and clinical endocrinology & diabetes* 122 (2014), Nr. 07, S. 384–386

[47] Kingma, Diederik P. ; Ba, Jimmy: Adam: A Method for Stochastic Optimization. In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015

[48] Kovatchev, Boris: A century of diabetes technology: signals, models, and artificial pancreas control. In: *Trends in Endocrinology & Metabolism* 30 (2019), Nr. 7, S. 432–444

[49] Kovatchev, Boris P. ; Breton, Marc ; Dalla Man, Chiara ; Cobelli, Claudio: *In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes.* 2009

[50] Kovatchev, Boris P. ; Cox, Daniel J. ; Gonder-Frederick, Linda A. ; Clarke, William: Symmetrization of the blood glucose measurement scale and its applications. In: *Diabetes Care* 20 (1997), Nr. 11, S. 1655–1658

[51] Kovatchev, Boris P. ; Renard, Eric ; Cobelli, Claudio ; Zisser, Howard C. ; Keith-Hynes, Patrick ; Anderson, Stacey M. ; Brown, Sue A. ; Chernavvsky, Daniel R. ; Breton, Marc D. ; Mize, Lloyd B. u. a.: Safety of outpatient closed-loop control: first randomized crossover trials of a wearable artificial pancreas. In: *Diabetes care* 37 (2014), Nr. 7, S. 1789–1796

[52] Lee, Hyunjin ; Bequette, B W.: A closed-loop artificial pancreas based on model predictive control: Human-friendly identification and automatic meal disturbance rejection. In: *Biomedical Signal Processing and Control* 4 (2009), Nr. 4, S. 347–354

[53] Lee, Seunghyun ; Kim, Jiwon ; Park, Sung W. ; Jin, Sang-Man ; Park, Sung-Min: Toward a fully automated artificial pancreas system Using a bioinspired reinforcement learning design: In silico validation. In: *IEEE Journal of Biomedical and Health Informatics* (2020)

[54] Leelarathna, Lalantha ; Choudhary, Pratik ; Wilmot, Emma G. ; Lumb, Alistair ; Street, Tim ; Kar, Partha ; Ng, Sze M.: Hybrid closed-loop therapy: Where are we in 2021? In: *Diabetes, Obesity and Metabolism* 23 (2021), Nr. 3, S. 655–660

[55] Lehmann, ED ; Deutsch, T ; Carson, ER ; Sönksen, PH: AIDA: an interactive diabetes advisor. In: *Computer methods and programs in biomedicine* 41 (1994), Nr. 3-4, S. 183–203

[56] Lewis, Dana: *OpenAPS. OpenAPS Data Commons The guide to DIY looping. 2014*. https://openaps.org/outcomes/data-commons/. – Accessed January 13, 2022.

[57] Little, Randie R. ; Sacks, David B.: HbA1c: how do we measure it and what does it mean? In: *Current Opinion in Endocrinology, Diabetes and Obesity* 16 (2009), Nr. 2, S. 113–118

[58] Ly, Trang T. ; Breton, Marc D. ; Keith-Hynes, Patrick ; De Salvo, Daniel ; Clinton, Paula ; Benassi, Kari ; Mize, Benton ; Chernavvsky, Daniel ; Place, Jéróme ; Wilson, Darrell M. u. a.: Overnight glucose control with an automated, unified safety system in children and adolescents with type 1 diabetes at diabetes camp. In: *Diabetes care* 37 (2014), Nr. 8, S. 2310–2316

[59] Ly, Trang T. ; Keenan, D B. ; Roy, Anirban ; Han, Jino ; Grosman, Benyamin ; Cantwell, Martin ; Kurtz, Natalie ; Eyben, Rie von ; Clinton, Paula ; Wilson, Darrell M. u. a.: Automated overnight closed-loop control using a proportional-integral-derivative algorithm with insulin feedback in children and adolescents with type 1 diabetes at diabetes camp. In: *Diabetes technology & therapeutics* 18 (2016), Nr. 6, S. 377–384

[60] Ly, Trang T. ; Roy, Anirban ; Grosman, Benyamin ; Shin, John ; Campbell, Alex ; Monirabbasi, Salman ; Liang, Bradley ; Eyben, Rie von ; Shanmugham, Satya ; Clinton, Paula u. a.: Day and night closed-loop control using the integrated Medtronic hybrid closed-loop system in type 1 diabetes at diabetes camp. In: *Diabetes Care* 38 (2015), Nr. 7, S. 1205–1211

[61] Maahs, David M. ; Buckingham, Bruce A. ; Castle, Jessica R. ; Cinar, Ali ; Damiano, Edward R. ; Dassau, Eyal ; DeVries, J H. ; Doyle, Francis J. ; Griffen, Steven C. ; Haidar, Ahmad u. a.: Outcome measures for artificial pancreas clinical trials: a consensus report. In: *Diabetes care* 39 (2016), Nr. 7, S. 1175–1179

[62] Magni, Lalo ; Raimondo, Davide M. ; Man, Chiara D. ; Breton, Marc ; Patek, Stephen ; De Nicolao, Giuseppe ; Cobelli, Claudio ; Kovatchev, Boris P.:

Evaluating the efficacy of closed-loop glucose regulation via control-variability grid analysis. In: *Journal of diabetes science and technology* 2 (2008), Nr. 4, S. 630–635

[63] Man, Chiara D. ; Micheletto, Francesco ; Lv, Dayu ; Breton, Marc ; Kovatchev, Boris ; Cobelli, Claudio: The UVA/PADOVA type 1 diabetes simulator: new features. In: *Journal of diabetes science and technology* 8 (2014), Nr. 1, S. 26–34

[64] Marling, Cindy ; Bunescu, Razvan: The OhioT1DM dataset for blood glucose level prediction: Update 2020. In: *CEUR workshop proceedings* Bd. 2675 NIH Public Access (Veranst.), 2020, S. 71

[65] Martinsson, John ; Schliep, Alexander ; Eliasson, Bjorn ; Meijner, Christian ; Persson, Simon ; Mogren, Olof: Automatic blood glucose prediction with confidence using recurrent neural networks. In: *KHD@ IJCAI*, 2018

[66] Martinsson, John ; Schliep, Alexander ; Eliasson, Björn ; Mogren, Olof: Blood glucose prediction with variance estimation using recurrent neural networks. In: *Journal of Healthcare Informatics Research* 4 (2020), Nr. 1, S. 1–18

[67] Mauseth, Richard ; Hirsch, Irl B. ; Bollyky, Jennifer ; Kircher, Robert ; Matheson, Don ; Sanda, Srinath ; Greenbaum, Carla: Use of a "fuzzy logic" controller in a closed-loop artificial pancreas. In: *Diabetes technology & therapeutics* 15 (2013), Nr. 8, S. 628–633

[68] Mauseth, Richard ; Wang, Youqing ; Dassau, Eyal ; Kircher Jr, Robert ; Matheson, Donald ; Zisser, Howard ; Jovanovič, Lois ; Doyle III, Francis J.: Proposed clinical application for tuning fuzzy logic controller of artificial pancreas utilizing a personalization factor. In: *Journal of diabetes science and technology* 4 (2010), Nr. 4, S. 913–922

[69] Naz, Huma ; Ahuja, Sachin: Deep learning approach for diabetes prediction using PIMA Indian dataset. In: *Journal of Diabetes & Metabolic Disorders* 19 (2020), Nr. 1, S. 391–403

[70] Ngo, Phuong D. ; Wei, Susan ; Holubová, Anna ; Muzik, Jan ; Godtliebsen, Fred: Control of blood glucose for type-1 diabetes by using reinforcement learning with feedforward algorithm. In: *Computational and mathematical methods in medicine* 2018 (2018)

[71] Nimri, Revital ; Muller, Ido ; Atlas, Eran ; Miller, Shahar ; Fogel, Aviel ; Bratina, Natasa ; Kordonouri, Olga ; Battelino, Tadej ; Danne, Thomas ; Phillip, Moshe: MD-Logic overnight control for 6 weeks of home use in patients with type 1 diabetes: randomized crossover trial. In: *Diabetes care* 37 (2014), Nr. 11, S. 3025–3032

[72] Pandit, Kaushik: Continuous glucose monitoring. In: *Indian journal of endocrinology and metabolism* 16 (2012), Nr. Suppl 2, S. S263

[73] Passino, Kevin M. ; Yurkovich, Stephen ; Reinfrank, Michael: *Fuzzy control*. Bd. 42. Citeseer, 1998

[74] Pedregosa, F. ; Varoquaux, G. ; Gramfort, A. ; Michel, V. ; Thirion, B. ; Grisel, O. ; Blondel, M. ; Prettenhofer, P. ; Weiss, R. ; Dubourg, V. ; Vanderplas, J. ; Passos, A. ; Cournapeau, D. ; Brucher, M. ; Perrot, M. ; Duchesnay, E.: Scikit-learn: Machine Learning in Python. In: *Journal of Machine Learning Research* 12 (2011), S. 2825–2830

[75] Peyser, Thomas ; Dassau, Eyal ; Breton, Marc ; Skyler, Jay S.: The artificial pancreas: current status and future prospects in the management of diabetes. In: *Annals of the New York Academy of Sciences* 1311 (2014), Nr. 1, S. 102–123

[76] Pickup, John ; Keen, Harry: Continuous subcutaneous insulin infusion at 25 years: evidence base for the expanding use of insulin pump therapy in type 1 diabetes. In: *Diabetes care* 25 (2002), Nr. 3, S. 593–598

[77] Pimentel, Eric C. ; Pimentel, Julio C.: A Novel Controller Architecture for Intelligent Artificial Pancreas. In: *2020 IEEE MIT Undergraduate Research Technology Conference (URTC)* IEEE (Veranst.), 2020, S. 1–4

[78] Polonsky, William: *Diabetes burnout: What to do when you can't take it anymore*. American Diabetes Association, 1999

[79] Revert, A ; Garelli, Fabricio ; Picó, Jesús ; De Battista, Hernán ; Rossetti, Paolo ; Vehí, Josep ; Bondia, Jorge: Safety auxiliary feedback element for the artificial pancreas in type 1 diabetes. In: *IEEE Transactions on Biomedical Engineering* 60 (2013), Nr. 8, S. 2113–2122

[80] Robertson, Gavin ; Lehmann, Eldon D. ; Sandham, William ; Hamilton, David: Blood glucose prediction using artificial neural networks trained with the

AIDA diabetes simulator: a proof-of-concept pilot study. In: *Journal of Electrical and Computer Engineering* 2011 (2011)

[81] Rossetti, Paolo ; Quiros, Carmen ; Moscardo, Vanessa ; Comas, Anna ; Giménez, Marga ; Ampudia-Blasco, F J. ; León, Fabián ; Montaser, Eslam ; Conget, Ignacio ; Bondia, Jorge u. a.: Closed-loop control of postprandial glycemia using an insulin-on-board limitation through continuous action on glucose target. In: *Diabetes technology & therapeutics* 19 (2017), Nr. 6, S. 355–362

[82] Sanz, Ricardo ; García, Pedro ; Díez, José-Luis ; Bondia, Jorge: Artificial Pancreas System With Unannounced Meals Based on a Disturbance Observer and Feedforward Compensation. In: *IEEE Transactions on Control Systems Technology* 29 (2020), Nr. 1, S. 454–460

[83] Schmitt, Andreas ; Reimer, André ; Kulzer, Bernhard ; Haak, Thomas ; Ehrmann, Dominic ; Hermanns, Norbert: How to assess diabetes distress: comparison of the Problem Areas in Diabetes Scale (PAID) and the Diabetes Distress Scale (DDS). In: *Diabetic Medicine* 33 (2016), Nr. 6, S. 835–843

[84] Seborg, Dale E. ; Edgar, Thomas F. ; Mellichamp, Duncan A. ; Doyle III, Francis J.: *Process dynamics and control*. John Wiley & Sons, 2016

[85] Sharifi, Amin ; De Bock, Martin I. ; Jayawardene, Dilshani ; Loh, Margaret M. ; Horsburgh, Jodie C. ; Berthold, Carolyn L. ; Paramalingam, Nirubasini ; Bach, Leon A. ; Colman, Peter G. ; Davis, Elizabeth A. u. a.: Glycemia, treatment satisfaction, cognition, and sleep quality in adults and adolescents with type 1 diabetes when using a closed-loop system overnight versus sensor-augmented pump with low-glucose suspend function: a randomized crossover study. In: *Diabetes technology & therapeutics* 18 (2016), Nr. 12, S. 772–783

[86] Shi, Dawei ; Deshpande, Sunil ; Dassau, Eyal ; Doyle III, Francis J.: Feedback control algorithms for automated glucose management in T1DM: the state of the art. In: *The Artificial Pancreas* (2019), S. 1–27

[87] Soru, Paola ; De Nicolao, Giuseppe ; Toffanin, Chiara ; Dalla Man, Chiara ; Cobelli, Claudio ; Magni, Lalo ; Consortium, AP@ H. u. a.: MPC based artificial pancreas: strategies for individualization and meal compensation. In: *Annual Reviews in Control* 36 (2012), Nr. 1, S. 118–128

[88] Steil, Garry M.: Algorithms for a closed-loop artificial pancreas: the case for proportional-integral-derivative control. In: *Journal of diabetes science and technology* 7 (2013), Nr. 6, S. 1621–1631

[89] Steil, Garry M. ; Rebrin, Kerstin ; Darwin, Christine ; Hariri, Farzam ; Saad, Mohammed F.: Feasibility of automating insulin delivery for the treatment of type 1 diabetes. In: *Diabetes* 55 (2006), Nr. 12, S. 3344–3350

[90] Sun, Qingnan ; Jankovic, Marko V. ; Budzinski, João ; Moore, Brett ; Diem, Peter ; Stettler, Christoph ; Mougiakakou, Stavroula G.: A dual mode adaptive basal-bolus advisor based on reinforcement learning. In: *IEEE journal of biomedical and health informatics* 23 (2018), Nr. 6, S. 2633–2641

[91] Sun, Xiaoyu ; Rashid, Mudassir M. ; Sevil, Mert ; Hobbs, Nicole ; Brandt, Rachel ; Askari, Mohammad-Reza ; Shahidehpour, Andrew ; Cinar, Ali: Prediction of Blood Glucose Levels for People with Type 1 Diabetes using Latent-Variable-based Model. In: *KDH@ ECAI*, 2020, S. 115–119

[92] Sutton, Richard S. ; Barto, Andrew G.: *Reinforcement learning: An introduction*. MIT press, 2018

[93] Tagougui, Sémah ; Taleb, Nadine ; Molvau, Joséphine ; Nguyen, Élisabeth ; Raffray, Marie ; Rabasa-Lhoret, Rémi: Artificial pancreas systems and physical activity in patients with type 1 diabetes: challenges, adopted approaches, and future perspectives. In: *Journal of diabetes science and technology* 13 (2019), Nr. 6, S. 1077–1090

[94] Tejedor, Miguel ; Woldaregay, Ashenafi Z. ; Godtliebsen, Fred: Reinforcement learning application in diabetes blood glucose control: A systematic review. In: *Artificial Intelligence in Medicine* 104 (2020), S. 101836

[95] Thabit, Hood ; Hovorka, Roman: Coming of age: the artificial pancreas for type 1 diabetes. In: *Diabetologia* 59 (2016), Nr. 9, S. 1795–1805

[96] Toffanin, Chiara ; Messori, Mirko ; Di Palma, Federico ; De Nicolao, Giuseppe ; Cobelli, Claudio ; Magni, Lalo: *Artificial pancreas: model predictive control design from clinical experience*. 2013

[97] *OpenAPS. Understanding Insulin on Board (IOB) Calculations. 2017.* https://openaps.readthedocs.io/en/latest/docs/While%

20You%20Wait%20For%20Gear/understanding-insulin-on-board-calculations.html. – Accessed January 23, 2022.

[98] Wang, Hongyu ; Chepulis, Lynne ; Paul, Ryan G. ; Mayo, Michael: Metaheuristic Optimization of Insulin Infusion Protocols Using Historical Data with Validation Using a Patient Simulator. In: *Vietnam Journal of Computer Science* 8 (2021), Nr. 02, S. 263–290

[99] Weisman, Alanna ; Bai, Johnny-Wei ; Cardinez, Marina ; Kramer, Caroline K. ; Perkins, Bruce A.: Effect of artificial pancreas systems on glycaemic control in patients with type 1 diabetes: a systematic review and meta-analysis of outpatient randomised controlled trials. In: *The lancet Diabetes & endocrinology* 5 (2017), Nr. 7, S. 501–512

[100] Woldaregay, Ashenafi Z. ; Årsand, Eirik ; Walderhaug, Ståle ; Albers, David ; Mamykina, Lena ; Botsis, Taxiarchis ; Hartvigsen, Gunnar: Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. In: *Artificial intelligence in medicine* 98 (2019), S. 109–134

[101] World Health Organization: *Global report on diabetes*. World Health Organization, 2016. – 83 p. S

[102] Xie, Jinyu: *Simglucose v0.2.1*. https://github.com/jxx123/simglucose. 2018. – Accessed June 28, 2021.

[103] Yamagata, Taku ; Ayobi, Amid ; O'Kane, Aisling A. ; Katz, Dmitri ; Stawarz, Katarzyna ; Marshall, Paul ; Flach, Peter A. ; Santos-Rodríguez, Raúl: Model-Based Reinforcement Learning for Type 1 Diabetes Blood Glucose Control. In: *Singular Problems for Healthcare Workshop at ECAI 2020*, 2020

[104] Zhu, Taiyu ; Li, Kezhi ; Herrero, Pau ; Georgiou, Pantelis: Basal Glucose Control in Type 1 Diabetes using Deep Reinforcement Learning: An In Silico Validation. In: *IEEE Journal of Biomedical and Health Informatics* (2020)

A Appendix

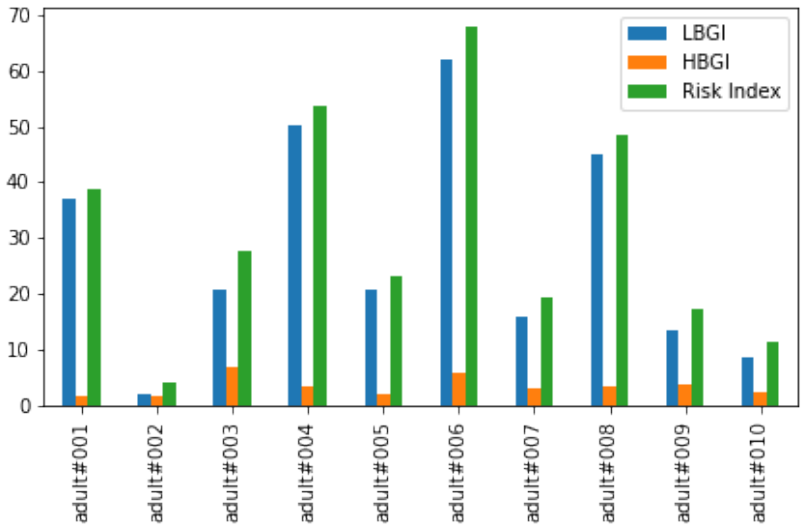

Figure A.1: Risk stats plot of the PID controller.

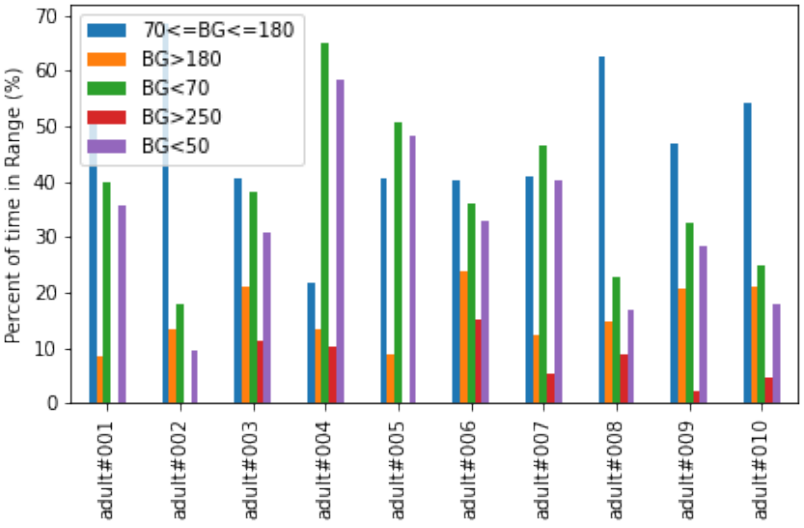

Figure A.2: Zone stats plot of the PID controller.

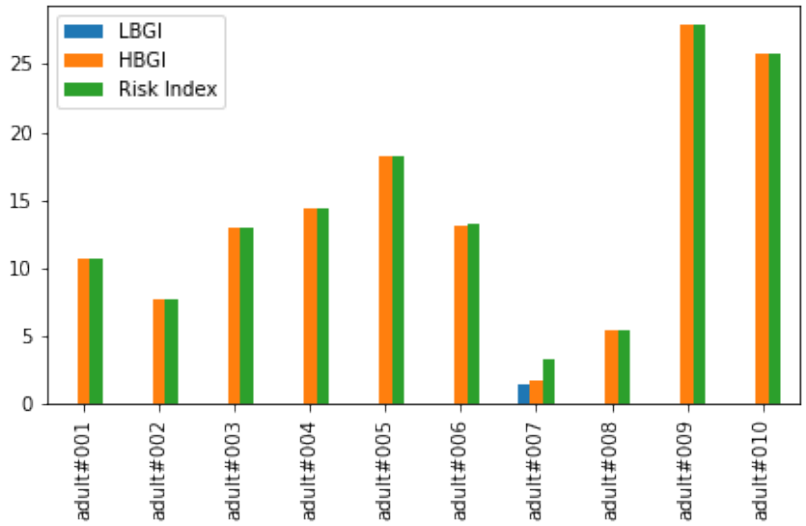

Figure A.3: Risk stats plot of baseline implementation.

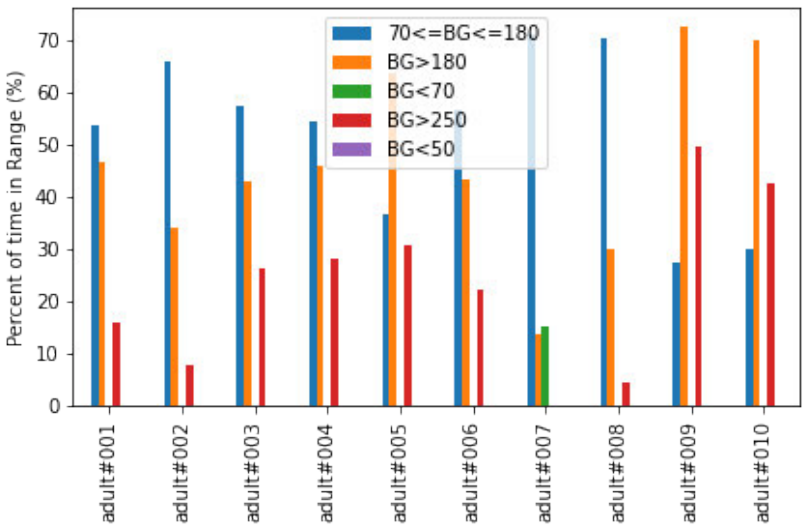

Figure A.4: Zone stats plot of baseline implementation.

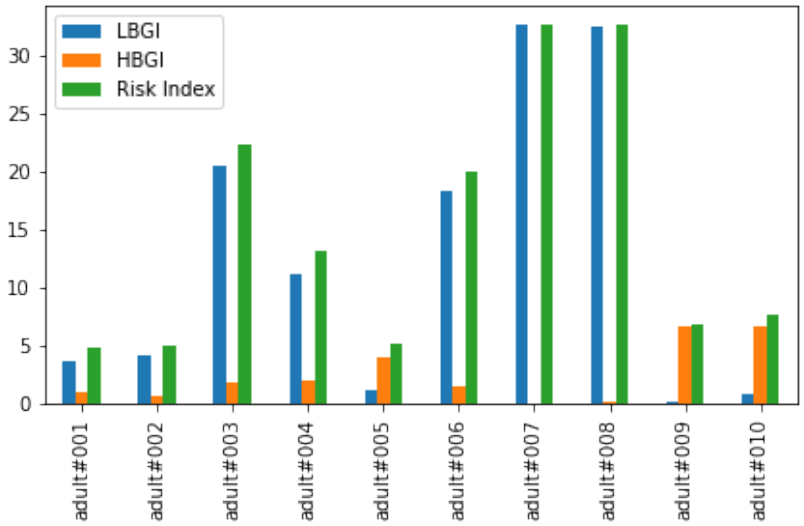

Figure A.5: Risk stats plot of applied scaling experiment.

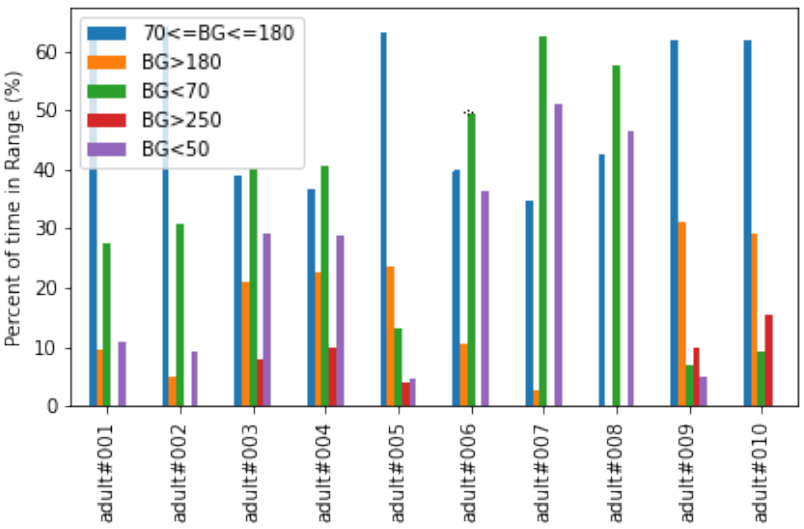

Figure A.6: Zone stats plot of applied scaling experiment.

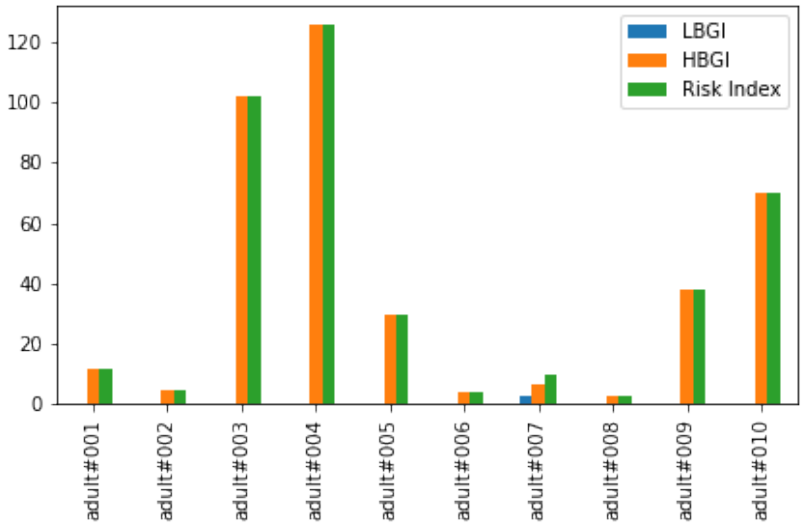

Figure A.7: Risk stats plot of bilinear IOB experiment.

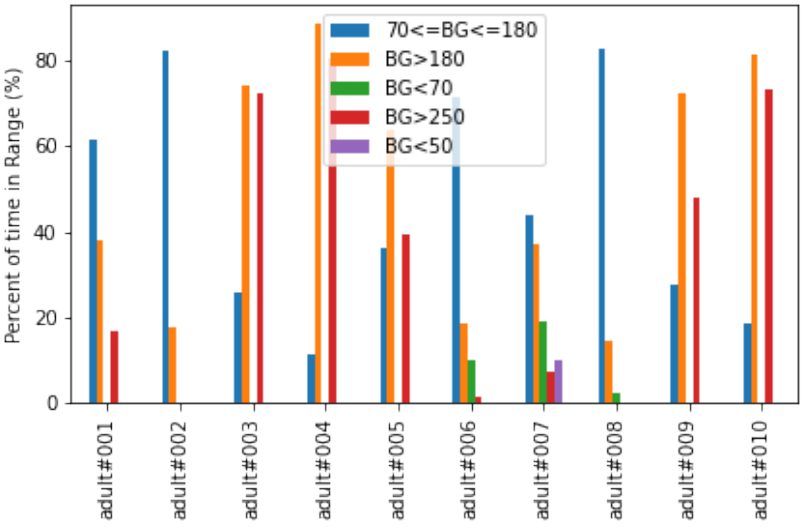

Figure A.8: Zone stats plot of bilinear IOB experiment.

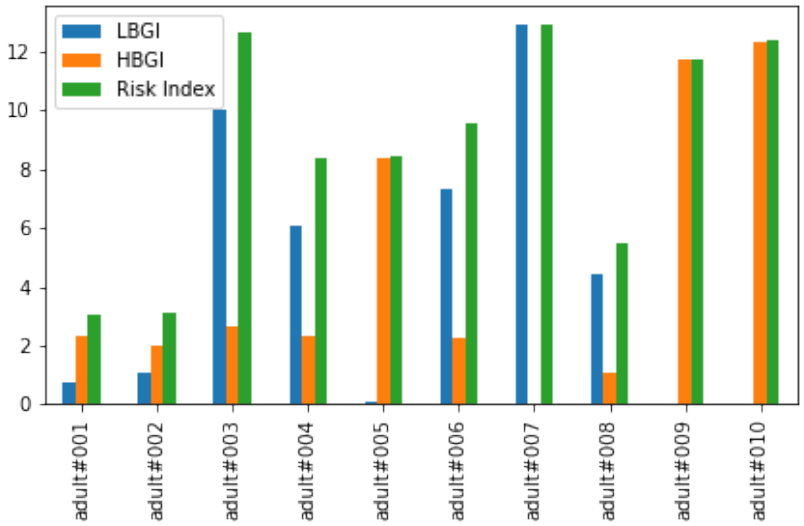

Figure A.9: Risk stats plot of exponential IOB experiment.

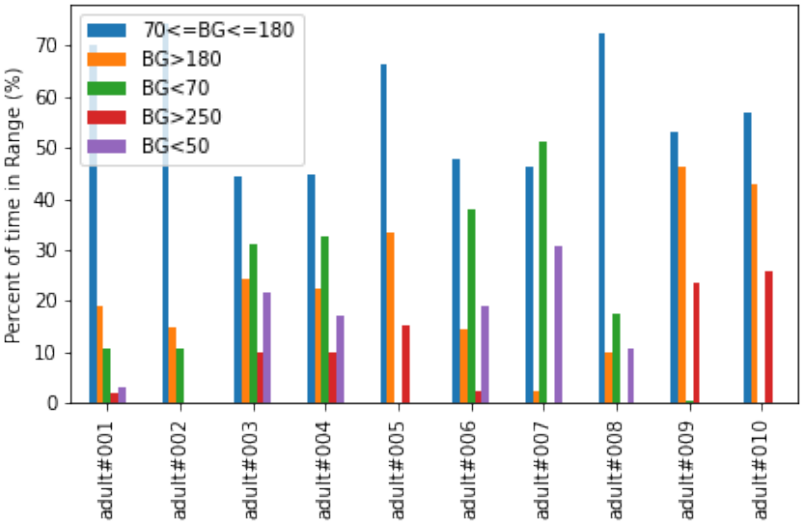

Figure A.10: Zone stats plot of exponential IOB experiment.

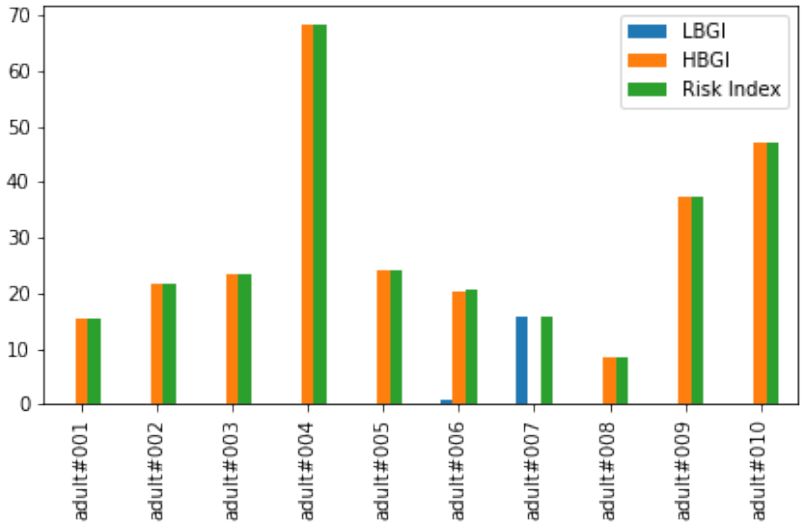

Figure A.11: Risk stats plot of exponential IOB (DIA 5h) experiment.

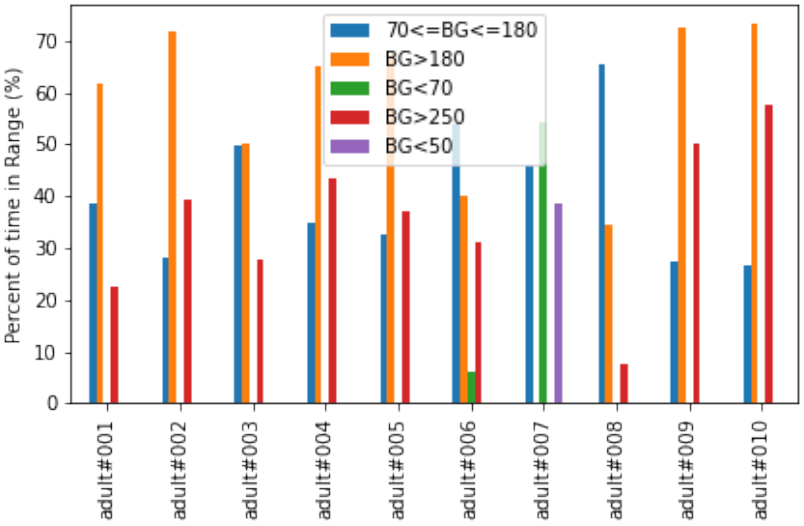

Figure A.12: Zone stats plot of exponential IOB (DIA 5h) experiment.

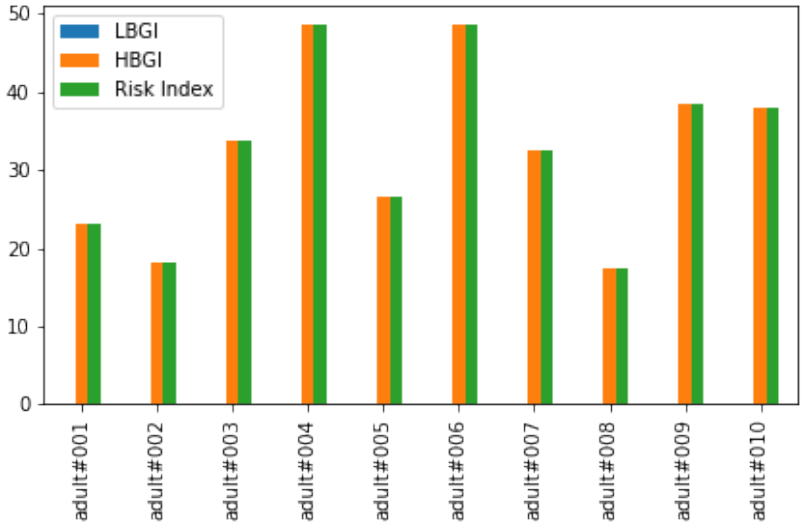

Figure A.13: Risk stats plot of refined action space experiment.

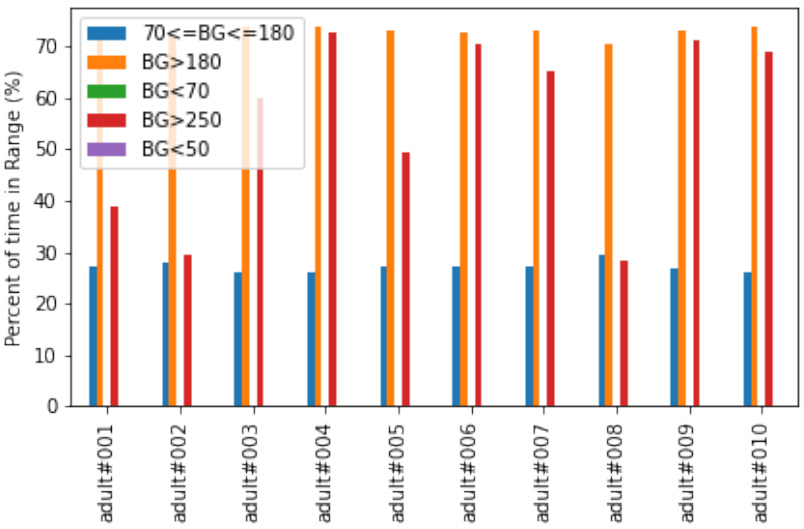

Figure A.14: Zone stats plot of refined action space experiment.

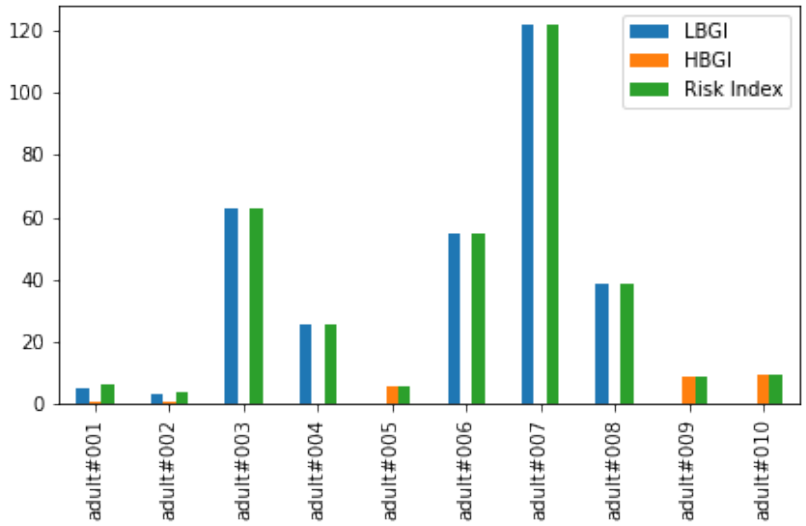

Figure A.15: Risk stats plot of refined action space (256x3) experiment.

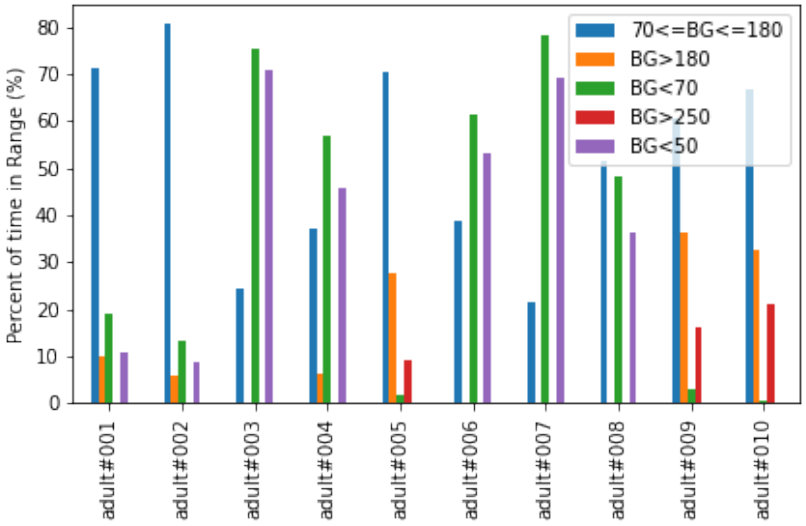

Figure A.16: Zone stats plot of refined action space (256x3) experiment.

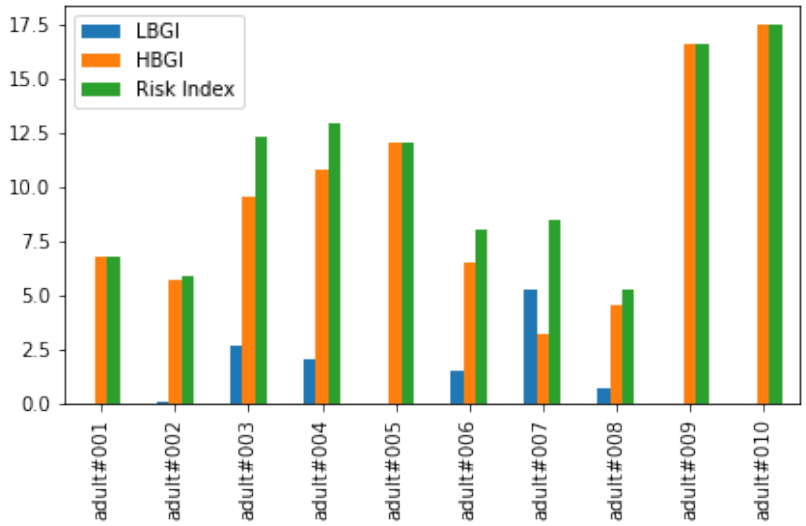

Figure A.17: Risk stats plot of refined action space (128x3) experiment.

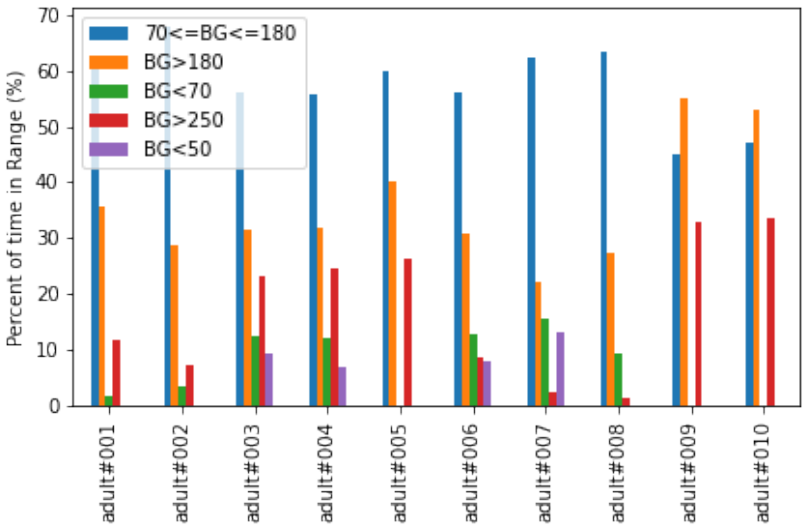

Figure A.18: Zone stats plot of refined action space (128x3) experiment.

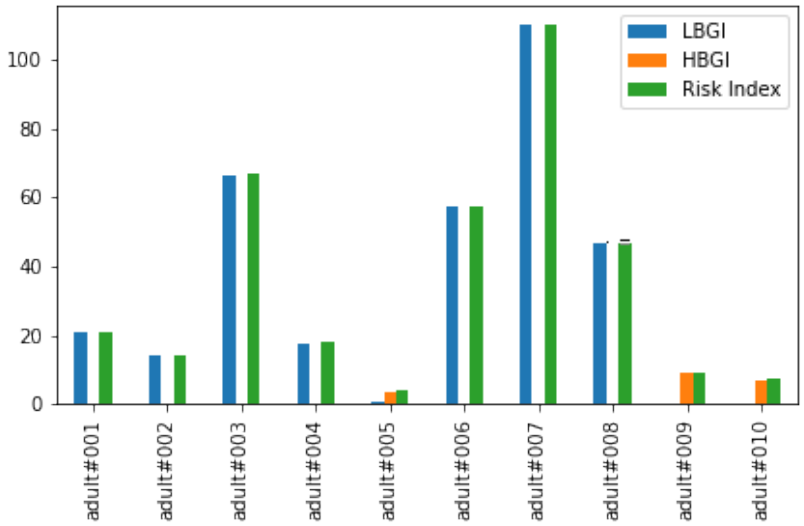

Figure A.19: Risk stats plot of different sequence length (16) experiment.

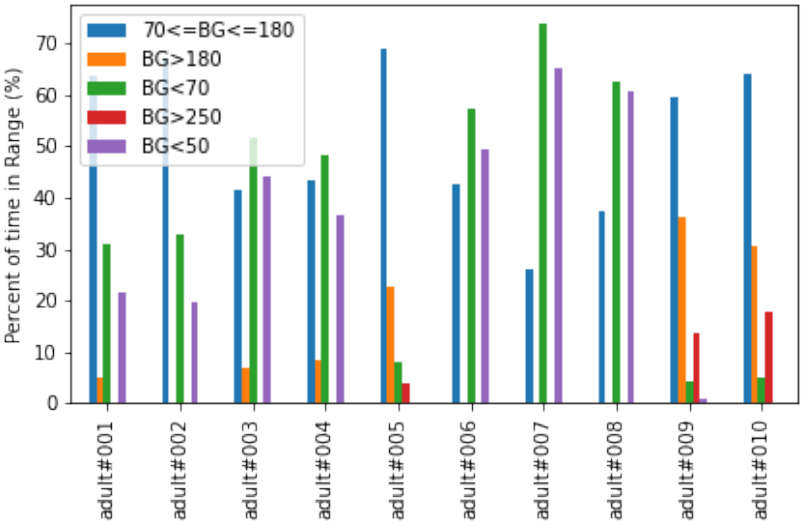

Figure A.20: Zone stats plot of different sequence length (16) experiment.

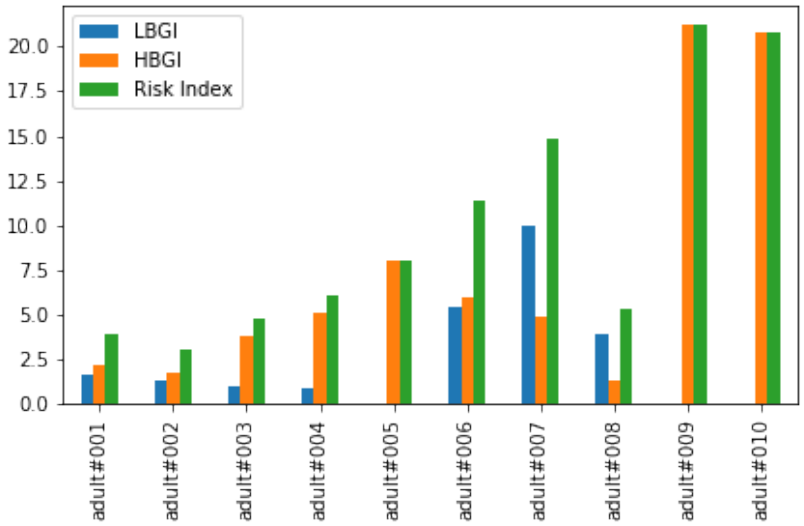

Figure A.21: Risk stats plot of different sequence length (8) experiment.

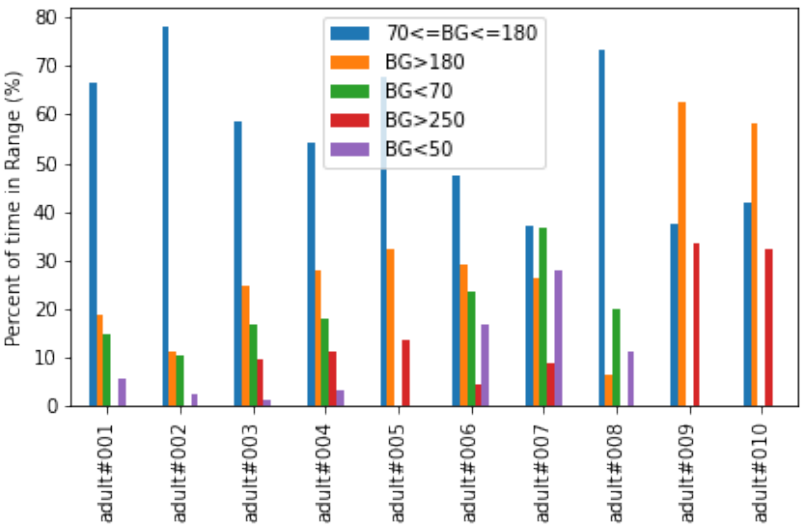

Figure A.22: Zone stats plot of different sequence length (8) experiment.

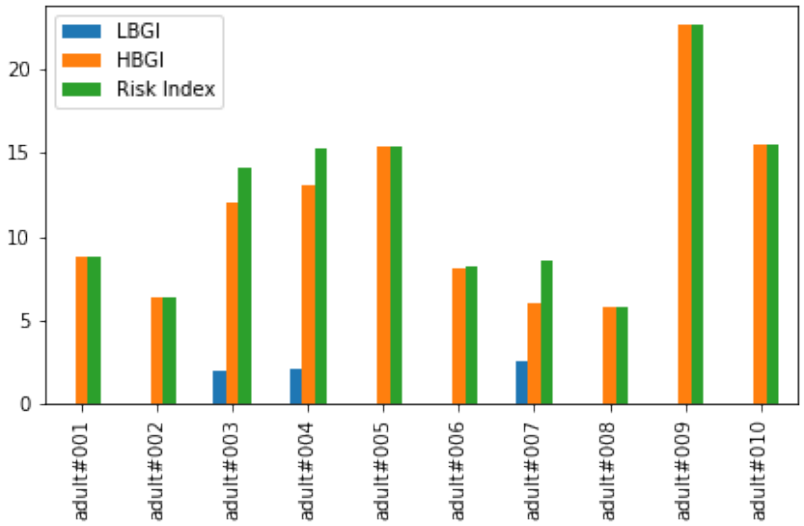

Figure A.23: Risk stats plot of different sequence length (64) experiment.

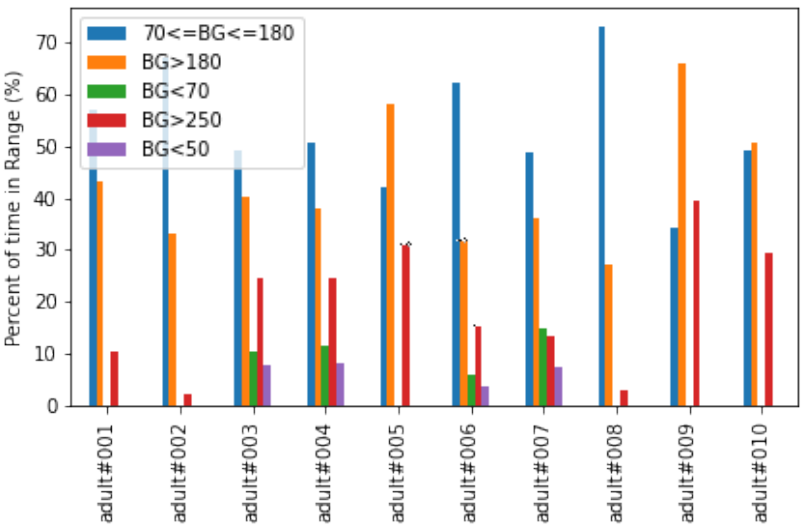

Figure A.24: Zone stats plot of different sequence length (64) experiment.

Erklärung zur selbstständigen Bearbeitung einer Abschlussarbeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

Ort	Datum	Unterschrift im Original