FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Assignment of bachelor's thesis

Title:	Analysis and prediction of blood glucose dynamics using Machine learning techniques
Student:	Ladislav Floriš
Supervisor:	Ing. Daniel Vašata, Ph.D.
Study program:	Informatics
Branch / specialization:	Knowledge Engineering
Department:	Department of Applied Mathematics
Validity:	until the end of summer semester 2022/2023

Instructions

Diabetes mellitus is a metabolic disorder that causes abnormal blood glucose (BG) regulation that might result in serious health complications if not properly managed. The current preferred treatment is primarily based on self-management of the disease, which means actively tracking BG levels and managing physical activity, diet, and insulin intake. The recent advancements in diabetes technologies and self-management applications have made it easier for patients to have more access to relevant data.

The objective of the thesis is to analyze the impact of physical activity, diet, and insulin intake on BG levels. Moreover, the second aim is to research and evaluate suitable machine learning techniques for the prediction of BG dynamics.

Electronically approved by Ing. Karel Klouda, Ph.D. on 8 February 2022 in Prague.

Bachelor's thesis

ANALYSIS AND PREDICTION OF BLOOD GLUCOSE DYNAMICS USING MACHINE LEARNING TECHNIQUES

Ladislav Floriš

Faculty of Information Technology
Department of Applied Mathematics
Supervisor: Ing. Vašata Daniel Ph.D.
May 12, 2022

Contents

List of Figures

List of Tables

I want to thank my supervisor Ing. Daniel Vašata Ph.D., for his support, patience, and all the great insights in our consultations. It truly was a pleasure. I also want to thank my family for their encouragement and for always believing in me. Further, I want to express gratitude to my partner for her support while working on this thesis as well as in everything I do. Finally, I thank all my friends from CTU for always being there during all the ups and downs of our lives and studies.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended. In accordance with Article 46(6) of the Act, I hereby grant a nonexclusive authorization (license) to utilize this thesis, including any and all computer programs incorporated therein or attached thereto and all corresponding documentation (hereinafter collectively referred to as the "Work"), to any and all persons that wish to utilize the Work. Such persons are entitled to use the Work in any way (including for-profit purposes) that does not detract from its value. This authorization is not limited in terms of time, location and quantity.

In Prague on May 12, 2022 .....

Abstrakt

Tato bakalářská práce se zabývá problematikou predikce hladiny glukózy v krvi u pacientů s diabetem typu 1. V naší práci nejprve analyzujeme změny koncentrace glukózy v krvi, a poté zkoumáme a vyhodnocujeme vhodné modely pro její predikci.

Zaměřili jsme se na modely založené na umělých neuronových sítích a support vector machines. Tyto modely byly experimentálně hodnoceny na 30minutovém, 1hodinovém a 2hodinovém predikčním horizontu. Data použitá v této práci byla shromážděna jedním pacientem po dobu 128 dnů a obsahují hodnoty krevní glukózy, dávky inzulínu, příjem sacharidů a fyzickou aktivitu.

Přesnost modelu byla hodnocena pomocí Root Mean Square Error (RMSE). K měření klinické přesnosti byla použita Clarke error grid analýza. Nejlepší dosažená RMSE byla 17,06 mg/dl, 24,32 mg/dl a 27,11 mg/dl pro 30minutový, 1hodinový a 2hodinový predikční horizont.

Naše výsledky ukazují, že je možné vyvinout modely pro predikci krevní glukózy použitelné v praxi. Na rozdíl od většiny prací věnujících se predikci krevní glukózy, jsme použili delší soubor dat, shromážděný po dobu 4 měsíců. Nakonec jsme dataset veřejně zpřístupnili pro další výzkum v této oblasti.

Klíčová slova diabetes typu 1, predikce hladiny krevní glukózy, prognóza časových řad, kontinuální monitorování glukózy, dynamika krevní glukózy

Abstract

This bachelor thesis tries to address the problem of predicting blood glucose (BG) levels of type 1 diabetes (T1D) patients. In our work, we first analyze BG dynamics and then research and evaluate suitable models for its prediction.

We focused on models based on artificial neural networks, and support vector machines. These models were experimentally evaluated on 30-minute, 1-hour, and 2-hours prediction horizons. The data used in this thesis was collected by one patient for 128 days in free-living conditions and contains BG levels, insulin doses, carbohydrate intake, and physical activity.

Model performance was assessed using Root Mean Square Error (RMSE). Clarke error grid analysis was used to measure clinical accuracy. The best RMSE achieved was 17,06 mg/dl, 24,32 mg/dl, and 27,11 mg/dl respectively for 30-minute, 1-hour, and 2-hours prediction horizons.

Our results show that it is possible to develop models for BG prediction which perform well in free-living conditions. Unlike most of the other papers in the academic literature on BG prediction, we used a longer dataset containing over 4 months' worth of data for a single patient. Lastly, we made this dataset publicly available for further research in this area.

Keywords type 1 diabetes, blood glucose prediction, time series forecasting, continuous glucose monitoring, blood glucose dynamics

List of Abbreviations

T1D	Type 1 diabetes
BG	Blood glucose
CGM	Continuous glucose monitoring
GI	Glycemic index
PH	Prediction horizon
ML	Machine learning
ANN	Artificial neural network
FFNN	Feed forward neural network
RNN	Recurrent neural network
LSTM	Long short-term memory
SVM	Support vector machine
RMSE	Root mean square error
CEGA	Clarke error grid analysis

Introduction

Type 1 diabetes (T1D) is an autoimmune disorder causing abnormal blood glucose (BG) levels due to the body's incapability to produce insulin. The current approach to treatment relies heavily on the patient's self-management, which means actively tracking glucose levels, injecting and dosing insulin, and managing diet and physical activity. It is not possible to cure T1D but patients can live without complications if they manage to keep their BG levels in the recommended range. Recent advancements in diabetes management technologies and wearable devices made the collection of relevant data more accessible.

We believe that machine learning (ML) techniques can utilize the available data and help patients with managing their BG levels. Mainly, we want to demonstrate that ML can be used for the prediction of BG. BG prediction is instrumental for the development of personalized decision systems, low and high BG alarms, and closed-loop systems[1], all of which aim to reduce the burden put on the patient and improve the T1D treatment.

The initial motivation for selecting this topic is that the author of this thesis lives with T1D as well and understands the complexity of managing the disease. There are tenths of research papers written on the topic of BG prediction but there are still very few real-world applications of these techniques. It is understandable that it takes time for commercial solutions to be developed as medical equipment has strict regulations and requires rigorous testing. The research often focuses on the development of closed-loop systems which do not only predict BG but are also capable of dosing insulin. Too low or especially too high insulin doses can result in serious complications and even death, meaning these systems must be safe and reliable. We believe that even a simpler system like a BG alarm which can alert the patient of a high or low glucose event has a vast impact on the patient's T1D management. If the model's inputs contain information about insulin and diet, it can even help the patient to make decisions regarding treatment, suggesting insulin intake or a portion of carbohydrates to consume in order to remain in a safe BG range.

Objectives and Goals

The goal of the thesis is to analyze the impact of insulin intake – both rapid and slow-acting insulin, diet (focused on carbohydrates), and physical activity on BG levels.

Secondly, we intend to research and evaluate machine learning techniques for the prediction of BG. Within the evaluation, the performance, and clinical accuracy of individual models is compared in order to identify the best-performing models for each prediction horizon (PH).

Thirdly, we want to create a dataset, which would be collected directly by the author of the theses. This dataset will contain glucose levels recorded by a Freestyle libre CGM sensor,

[1] A system capable of using the sensor glucose readings to calculate insulin dosage and send a signal to an insulin pump to administer insulin.

insulin, and carbohydrates intake notes – in form of an electronically maintained diary, and physical activity recorded by a Fitbit Charge fitness tracker. This data will be collected over a 4-month period in free-living conditions.

Thesis Overview

The thesis is structured into the 5 following chapters.

- Chapter 1 introduces T1D and factors relevant to the disease. Additionally, a theoretical model of BG prediction is introduced and research on the existing approaches for BG prediction is conducted.
- Chapter 2 describes individual data sources, data collection methodology, and characteristics of individual features.
- Chapter 3 analyzes BG dynamics and effects of various factors on the BG levels.
- Chapter 4 describes preprocessing and feature extraction. Furthermore, BG prediction models are introduced and their performance and clinical accuracy is evaluated.
- In Chapter 5 we summarize the achieved results and discuss potential improvements and applications.

Chapter 1
Theory

1.1 Type 1 Diabetes and BG Dynamics

We have put forward some fundamental facts about T1D in the introduction. This section dives deeper into the topic and describes the condition in more detail. Afterward, the problem of BG regulation and dynamics is introduced.

The main purpose of T1D treatment is to keep the BG in a safe range, by that we understand the observed range of a healthy person. The term for this safe range is **euglycemia**. The exact numbers used to classify euglycemia vary across the academic literature, but generally are around 60 mg/dl – 140 mg/dl or 3,3 – 7,8 mmol/l. This range is commonly used for diabetes tests [1].

Hypoglycemia is a condition when BG levels drop below 3,3 mmol/l or 60 mg/dl. Symptoms of hypoglycemia range from anxiety, sweating, and hunger, to neurological impairments, including behavioral changes, cognitive dysfunction, and in extreme cases seizures, and coma [2]. Generally, the lower the glucose, the higher the severity of symptoms. Patients' objective is to avoid hypoglycemia, as it impairs their function at best, and threatens their life at worst.

Hyperglycemia is a condition when BG is above 180 mg/dl or 10 mmol/l. Chronic levels of BG above this threshold can produce noticeable organ damage over time. Symptoms of hyperglycemia are usually benign like dry mouth and polyuria but can be more severe if hyperglycemia develops into ketoacidosis. The greatest danger of chronic hyperglycemia, which is frequent and prolonged incidents of hyperglycemia are the long-term effects, especially on the microvascular system. These effects may be life-threatening and include damage to the eye, kidneys, nerves, heart, and the peripheral vascular system [3].

1.1.1 Insulin

Insulin holds a significant role in T1D diabetes. Currently, most patients with T1D must decide on the appropriate doses of insulin to take, and/or rely on the doses recommended by their doctor. This decision-making places a big burden on the patients. Generally, there are two ways to approach this problem. Patients may decide to have a routine, with a consistent amount of carbohydrates consumed every day and stable insulin doses. The other option is to eat and function with fewer restrictions and without a routine but in that case, the patient must appropriately change the insulin doses as the situation requires. There is no single and exact equation that patients could use to calculate the appropriate insulin dose and it very often comes down to experience and trial and error. There are two means of administering insulin: insulin pumps and insulin pens.

Insulin pen is a reusable syringe with smart controls. Some pens have features like a

memory mechanism where the pen can display the time and dosage of the most recent insulin injection. Patients usually have a set of 2 pens, one containing rapid-acting insulin and the other a long-acting insulin.

Rapid-acting insulin (further referred to just as rapid insulin) is administered before and/or after food and its activity usually peaks after 2 hours after administration and ends after 6 hours. The exact times differ between insulin producers and types. Patients generally use rapid insulin at least 3 times a day (before main meals) but usually more often to administer corrective injections when BG rises above a safe range.

Long-acting insulin (further referred to just as long insulin) is usually injected before sleep and has a long but subtle activity. Some long insulins may only be effective for the night for a time period of around 10 hours and others may be active for as long as 24 hours. Long insulin from the Toujeo SoloStar pen, which is used by the patient whose data was collected as part of this thesis, has a 24 hours long activity.

Insulin pump is an electronic device connected at all times to the patient with a cannula. Insulin pumps usually have a user interface that allows convenient insulin administration and some additional advanced features. Insulin pumps tend to contain just one type of insulin and the long-lasting insulin is replaced by more frequent low doses called basal doses. A bolus dose is a bigger dose pumped to cover for food consumed and for spikes in BG which are to be corrected. Insulin pumps are the stepping stone for closed-loop systems as they have a mechanism for automatic insulin administration. However, for a closed-loop system to work, the insulin doses must be calculated in an automatic manner. For that reason, a precise predictive model has to be developed which would be capable of controlling the insulin pump doses.

The patient providing data for this thesis uses insulin pens, therefore our data contains records of rapid and long insulin.

1.1.2 BG Dynamics

Simply said, when studying BG dynamics, factors influence BG in two ways: they make BG rise or fall. For the purpose of BG dynamics, factors that do not have an influence on the BG levels are not examined.

The main factor that makes BG rise is carbohydrate consumption. After consumption, carbohydrates are broken down in the body into glucose, a very simple carbohydrate that the body uses as fuel for various physiological processes. The time between meal consumption and glucose availability in the blood varies between different foods. The length and rate of glucose release after food consumption for various foods is classified using Glycemic Index (GI) tables [4]. Some other factors that make BG rise are stress, anaerobic exercise, effects of glucagon – a hormone that stimulates the release of glucose stored in the liver, and others.

The main factor that makes BG fall is insulin. As previously stated, there are different insulin types and their activity varies, that is the onset and overall duration. These differences will translate to diverse effects on BG. Some other factors that make BG fall include aerobic exercise, alcohol consumption, and others.

It should also be noted that BG response can change depending on the daily insulin dose, as Davidson et al. demonstrated with adult patients using rapid insulin [5]. This effect can be summed up as follows: the higher the overall insulin dose taken daily by the patient, the lower the BG sensitivity to insulin. Furthermore, BG dynamics will vary between patients, for example, based on the body weight, as was again demonstrated by Davidson et al. [5].

1.2 BG Prediction

It was shown by Martin H Kroll [6] that BG variation includes a deterministic component and that one can describe it as a nonlinear oscillatory or chaotic system, that can be modeled.

If only past BG values are used for predictions, then for any time t the model inputs are the s previous BG measurements. The output is the predicted blood glucose $BG(t + PH)$, where the Prediction Horizon (PH) is how long into the future predictions are made. The goal is to find a function that takes a vector x containing past s BG measurements as an input and its output is the BG at time $t + PH$.

Since BG dynamics are influenced by factors like insulin, physical activity, and others, it makes sense to add these features to give the model more context. It is true that to a certain degree this information is already captured in the BG levels, but it may not provide the full picture. Therefore the prediction task is extended. The goal is to find a function $f(x_1, \ldots, x_n)$ which takes an ordered set of vectors as an input and its output is the BG at time $t + PH$. Each input vector x_i is associated with one of the features (BG, insulin, carbohydrates, ...). Each of the input vectors contains a finite number of elements and in the simple form, it will be the s successive previous measurements. However, greater input customizability is supported, by allowing the window of historical measurement to be variable in size between individual feature vectors. More than that, feature vectors that are transformations of the previous measurements are also permitted. The transformation may be the sum, mean, or other function applied to previous values.

It is expected that with a longer PH, the prediction performance will decrease. The uncertainty increases with the amount of time passed and so does the domain of possible BG values that come with a larger PH. It takes time for BG to change, and as PH increases, the BG has further opportunity to develop, thus the domain of possible BG values increases too. Figure 1.1 shows a representation of a prediction window with inputs being the normalized past BG measurements and true value (label) being the BG two steps (30 minutes) ahead.

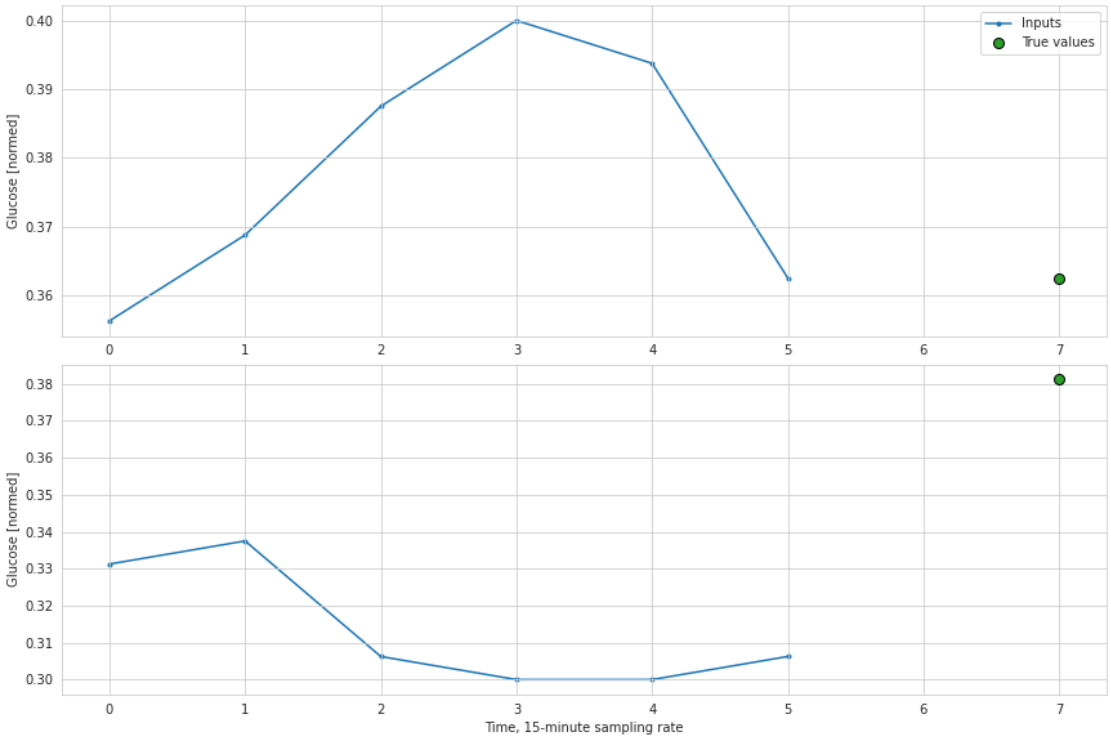

Figure 1.1 Example of a BG prediction window with inputs and true values

1.3 Existing Approaches

The starting point for the research of available literature on BG prediction is a comprehensive review [7] from Ashenafi Zebene Woldaregay et al. The review gives an insight into the types of models that were used, prediction horizons, and achieved performance. Further, there is information about study subjects and inputs used for the models. Figure 1.2 from the review [7] presents the Root Mean Square Error (RMSE) performance of some of the models' performances grouped by class of ML used.

Machine Learning	Validation	RMSE (mg/dl)(Mean/ Mean ±SD)				
		15 min	30 min	45 min	60 min	120 min
Recurrent Neural Network (RNN)	Random subsampling	2.52	7.56	15.12	23.76	
Feed Forward Neural Network (FFNN)	Random subsampling	2.70	7.56	14.94		
Hybrid (Feed Forward Neural Network Plus Linear Prediction Algorithm) along with Physiological Model	k-fold cross-validation		9.4			
Hybrid (Jump Neural Network along with Physiological Model)	k-fold cross-validation		16.6 ± 3.1			
Feed Forward Neural Network (FFNN)	10-fold cross-		13.31 ± 4.47		22.66 ± 6.86	37.62 ± 11.79
Self-organizing map (SOM)	validation		11.42 ± 2.33		19.58 ± 3.80	31.00 ± 6.07
A Neuro-Fuzzy Network with Wavelets as Activation Functions (WFNN)			15.22 ± 2.17		24.66 ± 3.39	39.59 ± 5.03
Hybrid (Compartmental model (MM)) and Self-Organizing Map (SOM) - Vector Quantization Method)	Hold-out		14.10 ± 4.57		23.19 ± 6.40	
Ensemble Approach Hybrid-Fusion (AR, Extreme Learning Machine, and Support Vector Regression- Kernel Function (Gaussian))	Hold-out		19.0 ± 0.3			
Feature based Feed-Forward Neural Network (FFNN)	Random subsampling		10.00	15.00	20.00	
Hybrid (Generic Physiological Model & Support Vector Regression- Gaussian kernel)	Random subsampling		22.6		35.8	
Feed-Forward Neural Network model using Guardian CGM	Hold-out	9.74 ± 2.71	17.45 ± 5.44	25.08 ± 8.73		
Feed-Forward Neural Network Model using FreeStyle Navigator CGM		10.38 ± 3.15	19.51 ± 5.53	29.07 ± 6.77		
Hybrid-Fused (ARX and Elman simple Recurrent Neural Network) for prediction and Extreme Learning Machine for correction	Hold-out	8.9 ± 1.70	18.9 ± 4.60	21.6 ± 4.39		
Single hidden layer Feedforward Neural Networks - (kernel RLS, Gaussian kernel) - Extreme Learning Machine	10-fold cross-validation		6.1 ± 1.6			
Support Vector Regression (SVR—RF)	10-fold cross-		5.7 ± 1.5		6.4 ± 2.1	
Support Vector Regression (SVR—RRF)	validation		5.9 ± 1.4		6.8 ± 2.0	
Gaussian Processes (GP—RF)	10-fold cross-		5.6 ± 1.7		6.3 ± 2.6	
Gaussian Processes (GP—RRF)	validation		5.9 ± 1.6		6.8 ± 2.9	
Hybrid- (Random Forests Regression technique & Compartmental Model)	10-fold cross-validation	6.60	8.15		9.25	10.83
Hybrid- (Support Vector Machines- Linear kernel and Compartmental Model)	v-fold cross validation	9.28	15.59		24.06	31.24
Hybrid- (Support Vector Machines- Linear kernel and Compartmental Model)	Random subsampling	9.1	14.8		22.4	28.2
Hybrid- (Support Vector Machines- Linear kernel and Compartmental Model)	v-fold cross validation	9.51 ± 2.39	16.02 ± 3.55		24.81 ± 4.74	36.15 ± 9.70
Hybrid- (Support Vector Machines- Linear kernel and Compartmental Model)	10-fold cross-validation	5.21	6.03		7.14	7.62
Hybrid-Fused (Autoregressive Model with output correction – cARX, & Recurrent Neural Network – RNN)-Data fusion (Genetic Algorithms (GA), & Genetic Programming (GP))	Hold-out	11.9	18.9	26.1		
Hybrid (Genetic Programming - Grammatical Evolution & Physiological model)	Hold-out				5.12	
Hybrid-(Compartmental model & Support Vector Regression- linear kernel) using Physical Activity – Sensor Data as Input	Leave-one-day out	11.13	18.84		28.79	46.7
Hybrid-(Compartmental model & Support Vector Regression- linear kernel) using Exercise Modelling		10.84	17.92		27.5	43.34
Recurrent Neural Network (RNN)	Hold-out		21.4		38.0	
Hybrid- (Compartmental Models (CMs) and a Self-Organizing Map (SOM) - Vector Quantization Method)	Hold-out		14.10 ± 4.57		23.19 ± 6.40	
Support Vector Regression (SVR) based on Differential Evolution (DE) Algorithms	Hold-out	9.44	10.78	11.82	12.95	
Feed Forward Artificial Neural Networks (FFNN)	Hold-out	6.43	7.45	8.13	9.03	

Figure 1.2 Literature review of predictive performance (RMSE) based on the class of ML employed, validation strategies, and PH considered on real subjects [7]

Models based on artificial neural networks (ANN) are the most common, followed by hybrid approaches and support vector machines (SVM), with most of the papers being published within the last ten years [7]. However, the topic of BG prediction is not completely recent. One of the first investigations was done already in 1999 by Bremer and Gough [8]. The authors showed that BG values in the near future could be predicted based only on previous BG values.

There are 2 publications that stand out when examining the prediction performance achieved in Figure 1.2, it is the ANN-based model proposed by Jaouher Ben Ali et al. [9] and the SVM-based model from Georga et al. [10].

What attracts attention to the SVM model proposed by Georga et al. [10] is its RMSE prediction performance of 7.62 mg/dl in a PH of 2 hours, which substantially exceeds the other models reviewed. The inputs used are the subcutaneous glucose profile, the plasma insulin concentration, the appearance of the meal-derived glucose in the systemic circulation, and the energy spent during physical activities. Data from 27 T1D patients collected in free-living conditions was used. They have proven that the availability of multi-variable data can increase prediction performance.

The ANN model by Jaouher Ben Ali et al. [9] evaluation has the longest PH of 1 hour, and even though their performance is worse than the SVM model by Georga et al., they used previous BG only as inputs to the model. The RMSE prediction performance for 15, 30, 45, and 60 minute PH is 6,43 mg/dl; 7,45 mg/dl; 8,13 mg/dl; and 9,03 mg/dl respectively, whereas SVM performance for the same prediction horizons is 5,21 mg/dl; 6,03 mg/dl; 7,14 mg/dl; and 7,62

mg/dl respectively. Therefore, the performance is surprisingly high given that the ANN model had considerably more limited information for predictions.

It has to be pointed out that it is problematic to compare prediction performance between the published papers as the model evaluation generally is not performed on the same dataset. This is problematic because some patients' BG dynamics may be easier to predict because they follow the same routine every day and their BG has a regular pattern. That however may not apply to other patients whose BG is more irregular due to a less predictable lifestyle. Additionally, the prediction methodology may differ. For example, for any time t and PH Georga et al. [10] dropped out of their dataset any windows where there was an event (i.e., food intake, insulin intake, moderate or intense exercise) at the time interval $[t, t + PH]$. This dropout of training instances makes sense as they do not represent a rational mapping between the input and the output [10]. However, it also makes the data less noisy and it is harder to make predictions if such dropout is not used. The increased difficulty comes from the uncertainty of future events influencing BG which must be accounted for.

Chapter 2
Dataset

In order to have as much freedom and control over the dataset as possible, we have decided to use data captured directly by the thesis author. The second reason for doing so, is that very few of the published articles include data that was used to perform the experiments. The downside of this strategy is that our evaluation can only capture the performance of the model on one patient. All data was captured in free-living conditions and with consumer-grade hardware. Data collected in a lab environment probably will not capture all the complex dynamics seen in free-living conditions. Thus, it can be argued, that performance evaluation on data from free-living conditions is more significant.

The data was collected from 1. 12. 2021 to 8. 4. 2022, which adds up to 128 days worth of data.

Due to the difficulty of finding an appropriate dataset online, we decided to make the dataset collected as part of this thesis available on Kaggle [11] [12]. Kaggle allows its users to share datasets with each other for research and development of ML models.

Our dataset has three sources of data. It is the data captured by a CGM sensor which is read and stored by a glucose reader. Data from an electronic diary containing information about insulin intake, carbohydrate intake, an estimate of food glycemic index, and some additional features which were not used in this thesis. The last data source is a Fitbit fitness tracker which contains a wide range of features, for the purposes of this theses, heart rate, distance, and burned calories are used.

2.1 Glucose Reader

The BG data was collected using a CGM developed by ABBOTT Laboratories called FreeStyle Libre. A more accurate term to use is flash-glucose-monitoring (FGM) because Freestyle Libre sensors do not stream the collected data. The sensor must be read with a reader that comes with it or with a phone with NFC using the Librelink app [13]. For the purposes of this thesis, it can be regarded as if it was a CGM since it still provides a continuous glucose history. The difference between FGM and CGM is mainly in the real-time streaming abilities which are not relevant for our purposes.

FreeStyle Libre is implanted under the skin and has to be replaced every 14 days. The sampling rate of the sensor is 15 minutes. The sensor stores an 8-hour glucose history. Any data written more than 8 hours ago gets overwritten. Thus, the reader must read the data from the sensor at least every 8 hours in order to prevent any data loss. The reader is also capable of displaying charts of blood glucose, trends, and some other useful information. It is possible to export data from the reader when it is connected to a computer. The reader has a limited

storage capacity, so exports must be done regularly. We have exported data approximately every 2 weeks, which is sufficient to prevent any data loss from the reader.

BG in our dataset is measured in mmol/l units, which is the number of molecules of a substance within a specified volume. In this case the amount of glucose within 1 liter. In the published literature on BG prediction, it is common to measure BG in mg/dl which gives the concentration by the ratio of weight to volume, in this case, milligrams of glucose per decilitre. In order to convert from mmol/l to mg/dl, the BG values must be multiplied by a conversion factor of 18,016. The molecular weight of glucose is 180,16 g/mol as per the periodic table. This number is then divided by 10 because of the change from liters to decilitres, and thus the 18,016 conversion factor is constructed.

It is essential to mention that CGM sensors actually measure glucose in the interstitial fluid and not in the blood. There is a lag between BG and interstitial glucose which ranges from 5 to 10 minutes according to an article from diabetes journals [14] and according to a study performed on an older FreeStyle Navigator sensor, it can go as high as 16,8 minutes in the extremes [15]. This lag does not influence the prediction task but has significance for the patient. It is mainly important in the events of rapidly rising or falling BG. When BG moves at a high rate, it can in a short time get to the hypoglycemic or hyperglycemic range. The patient may be feeling symptoms of hyperglycemia or hypoglycemia even though the sensor is reporting BG values in a safe range.

2.2 Electronic Diary

The electronic diary was at first realized using the FreeStyle Libre reader as it allows patients to add notes about insulin and carbohydrate intake. However, we moved to the mySugr mobile app [16] on 9. 1. 2022 because it supported a broader range of features to capture and was easier to work with as well.

There are four main features captured in the electronic diary, and for each record, in the diary, there is a timestamp.

- Rapid insulin from Fiasp FlexTouch pen was used. Each rapid insulin record contains a dosage in units. There are 100 units in 1 ml of Fiasp insulin.
- Long insulin from Toujeo SoloStar pen was used. Each record for long insulin also contains a dosage in units. Toujeo measures units differently and there are 300 units in 1 ml.
- Carbohydrates, as an estimate of the grams of carbohydrates consumed, are recorded by the patient for each meal. The record is noted approximately at the beginning of the meal consumption.

  The patient estimates the carbohydrate load of food that he prepares himself by weighting individual components of the food which contain carbohydrates and then multiplying the weight by the percentage of carbohydrates contained in each component. The patient relies on nutritional information written on the packaging and/or in the database from the app Kalorické tabulky [17]. It is necessary to point out that the accuracy of these records may vary based on the accuracy of nutritional information as well as based on the food, where carbohydrate load is harder to determine.
- Glycemic Index (GI) is written down by the patient as an estimate of the consumed meal GI, for each meal that contains carbohydrates. The GI feature is categorical and is mostly grounded in the table seen in Figure 2.1 from the International tables of glycemic index [4].

Western refined foods			Unrefined traditional foods		
Food	Glycemic index	Glycemic load	Food	Glycemic index	Glycemic load
Glucose	97	96.8	Parsnips	97	19.5
Rice Krispie cereal	88	77.3	Baked potato	85	18.4
Cornflakes	84	72.7	Boiled millet	71	16.8
Lifesavers	70	67.9	Boiled broad beans	79	15.5
Rice cakes	82	66.9	Boiled couscous	65	15.1
Table sugar (sucrose)	65	64.9	Boiled sweet potato	54	13.1
Shredded wheat cereal	69	57.0	Boiled brown rice	55	12.6
Graham crackers	74	56.8	Banana	53	12.1
Grapenuts cereal	67	54.3	Boiled yam	51	11.5
Cheerio cereal	74	54.2	Boiled garbanzo beans	33	9.0
Rye crispbread	65	53.4	Pineapple	66	8.2
Vanilla wafers	77	49.7	Grapes	43	7.7
Corn chips	73	46.3	Kiwi fruit	52	7.4
Mars bar	68	42.2	Carrots	71	7.2
Stone wheat thins	67	41.9	Boiled peas	48	6.8
Shortbread cookies	64	41.9	Boiled beets	64	6.3
Granola bar	61	39.3	Boiled kidney beans	27	6.2
Angel food cake	67	38.7	Apple	39	6.0
Bagel	72	38.4	Boiled lentils	29	5.8
Doughnuts	76	37.8	Pear	36	5.4
White bread	70	34.7	Watermelon	72	5.2
Waffles	76	34.2	Orange	43	5.1
All bran cereal	42	32.5	Cherries	22	3.7
Whole wheat bread	69	31.8	Peach	28	3.1
Fructose	23	22.9	Peanuts	14	2.6

Figure 2.1 Glycemic Index overview for common food

The GI feature can contain 4 different categorical values: Low, Medium, High, and Very High. Their relation to the GI is following:

1. Low: 0–30 GI
2. Medium: 30–60 GI
3. High: 60–90 GI
4. Very High: 90–100 GI

The GI estimation is problematic, especially for meals containing many components, and therefore this feature may not be very reliable. However, it may be beneficial for estimating the rate of change of BG after meals. Higher GI should result in a more rapid impact on BG whereas with lower GI a more gradual impact on BG is expected. The GI data acquisition began later on in the data collection period – on 9. 1. 2022.

2.3 Fitbit Fitness Tracker

Physiological data was collected using a Fitbit Charge 4 fitness tracker. Before the data collection began, we went over the available literature on the estimation of accuracy of consumer-grade fitness trackers and smartwatches. In a meta-analysis published in the British Journal of Sports Medicine, [18] a wide range of consumer-grade fitness trackers were evaluated in their ability to estimate energy expenditure. Energy expenditure is usually calculated based on heart rate, distance, and skin temperature. The Fitbit Charge device was chosen due to its affordability and data export capabilities. Figure 2.2 taken from the meta-analysis shows, that Fitbit devices do not significantly overestimate or underestimate energy expenditure.

There is a variety of data in the export provided by Fitbit like heart rate, distance, steps, sleep length and quality, respiratory rate, calories burned, and more.

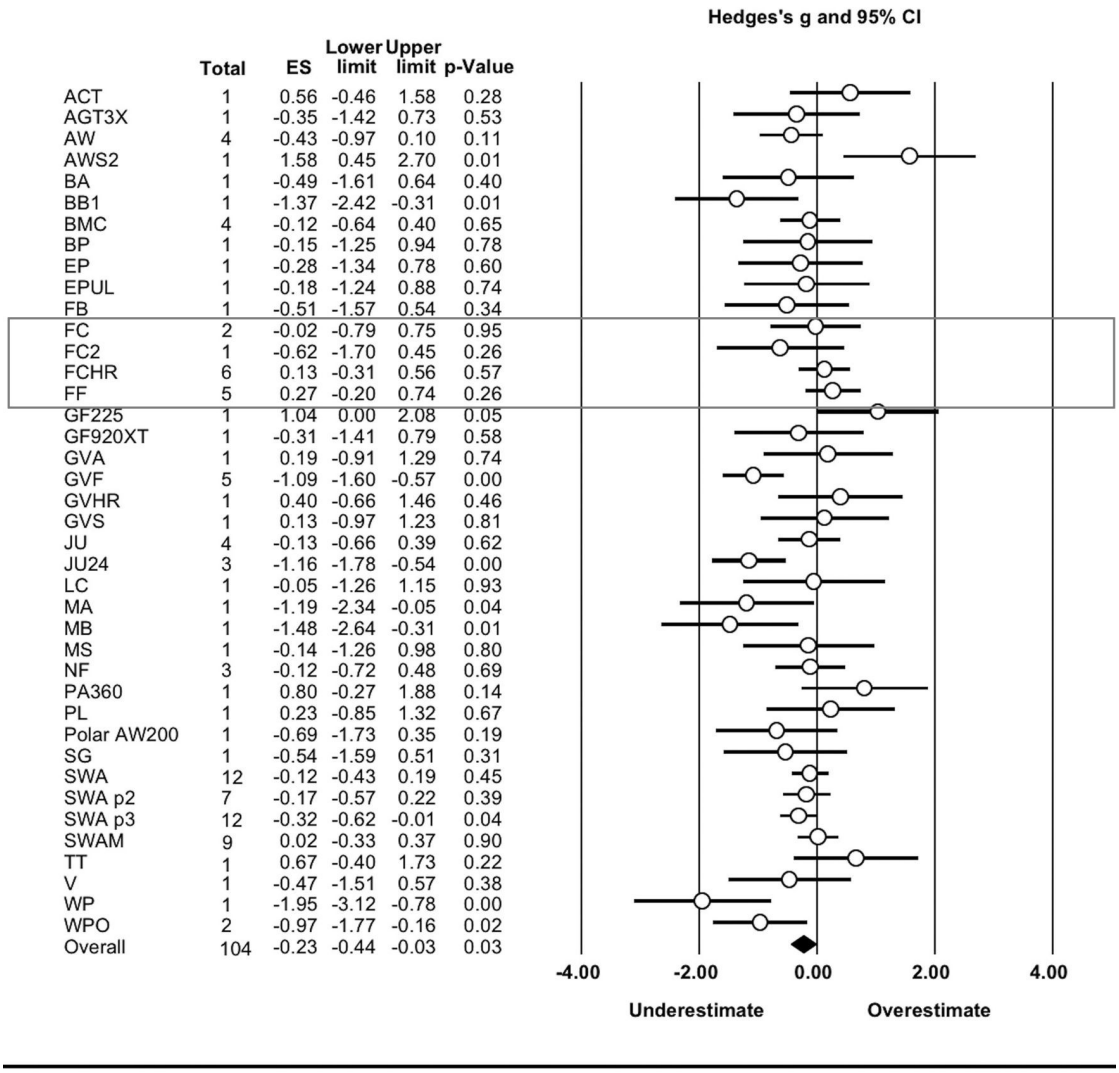

Figure 2.2 Overall energy expenditure accuracy comparison between fitness trackers, Fitbit devices highlighted [18]

In this thesis, only heart rate, distance, and calories are considered, as these sorts of metrics are the most readily available in fitness trackers.

- Heart rate is measured in beats-per-minute and the tracker produces a heart rate estimate approximately every 5 seconds.
- Distance is measured in centimeters covered and has a minutely sampling rate.
- Calories are measured in kcal burned with a 1-minute sampling rate.

Chapter 3
Analysis

In this chapter, an exploratory data analysis of the collected data is performed in order to analyze the impact of factors like insulin, carbohydrates, physical activity, and others on the BG levels. The data manipulation and visualizations were created in JupyterLab computational environment [19]. Pandas library [20] is used for data manipulation and matplotlib [21] and seaborn [22] libraries for visualizations.

3.1 Basic Exploratory Analysis and BG Behavior

In Table 3.1 basic descriptive statistics are shown, like the number of samples, mean, standard deviation, minimum and maximum, and 25, 50, and 75 percentile values. This presents an initial insight into the feature properties. The standard deviation indicates the carbohydrates consumed are spread out the most. Insulin doses, especially for long insulin, are less dispersed, meaning the patient does not tend to change the doses very often.

Table 3.1 Descriptive statistics of BG, Rapid Insulin, Long Insulin, and Carbohydrates

	BG [mg/dl]	Rapid Insulin [units]	Carbohydrates [grams]	Long Insulin [units]
count	11970	640	1372	126
mean	137,69	5,01	16,01	12,55
std	51,06	2,84	13,27	2,24
min	50,44	1	2	5
max	340,5	10	70	15
25%	99,09	2	7	11
50%	131,52	5	11	13
75%	169,35	8	20	15

To get a better idea of the underlying distributions, BG and carbohydrates distributions are shown in Figure 3.1. It is visible that carbohydrates distribution has a longer tail than BG.

Long-term BG behavior is shown in Figure 3.2 with important thresholds highlighted. In order to get a better insight into the causes of BG changes, a single day of BG behavior is displayed in Figure 3.3 with a scatter-plot on top indicating events of meals and insulin injections. Each scatter point also contains the number of grams of carbohydrate or insulin dosage. All scatter points are min-max normalized.

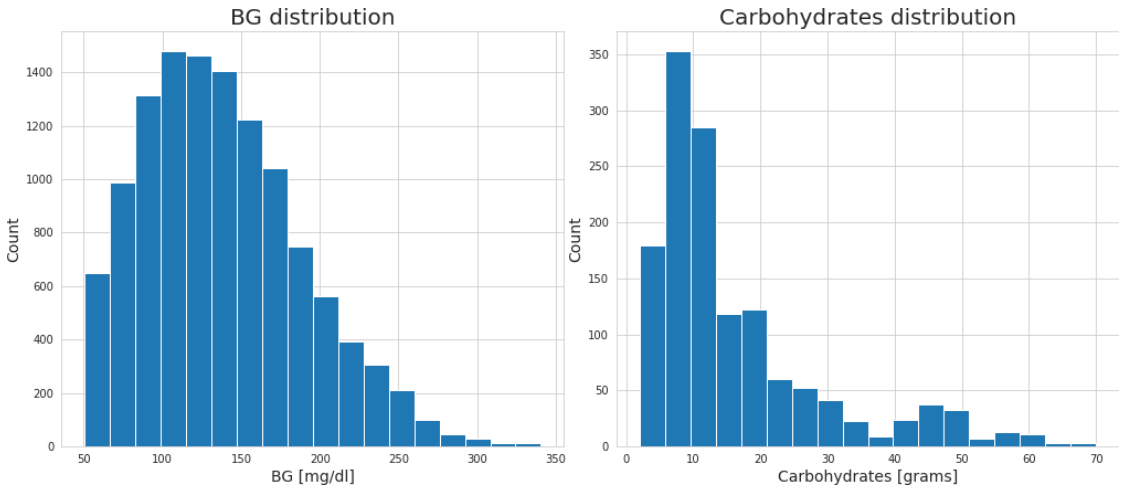

Figure 3.1 Distribution of BG levels and carbohydrates consumed

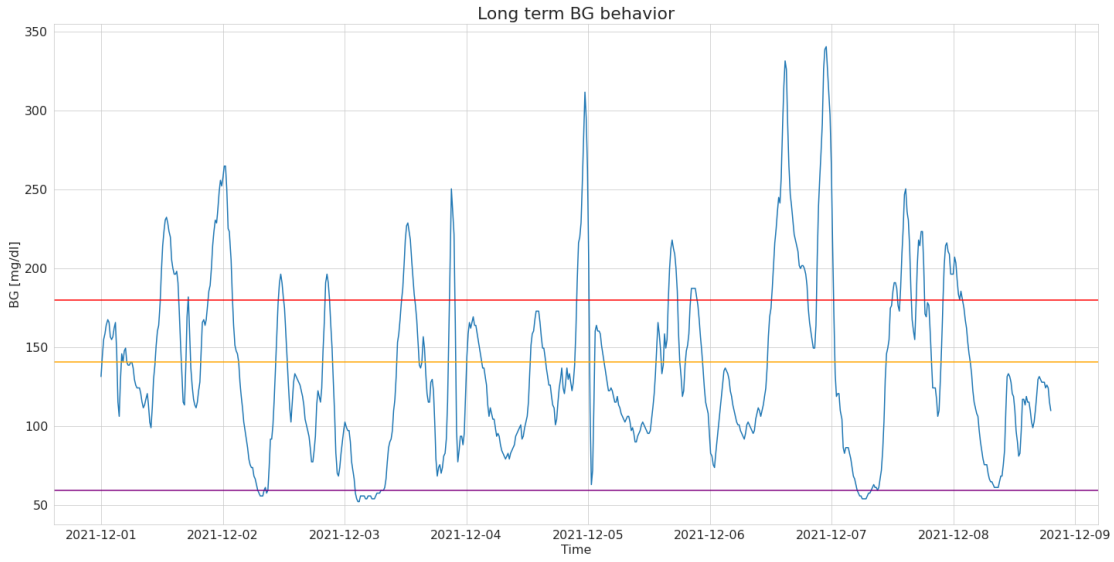

Figure 3.2 Long-term BG behavior with purple line threshold for hypoglycemia, orange for euglycemia, and red for hyperglycemia

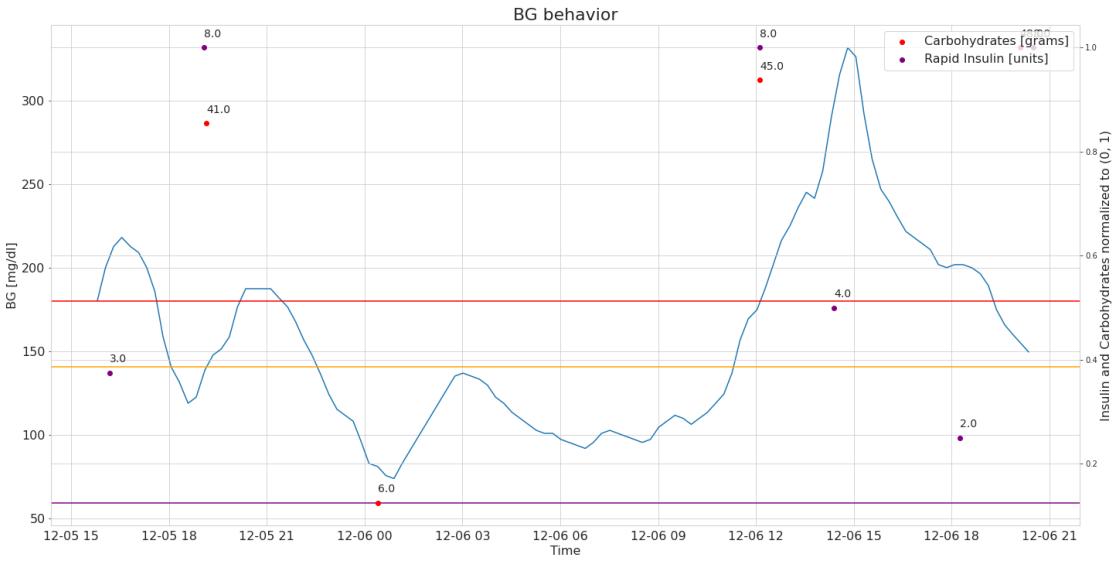

Figure 3.3 Short-term BG behavior labeled with insulin and carbohydrates

3.2 Blood Glucose Sensitivity to Insulin & Carbohydrates

In this section, BG changes are explored in response to insulin injections and carbohydrate consumption. Changes are investigated in 30-minute and 1-hour windows, $t+\frac{1}{2}$hour and $t+1$hour, with an insulin or carbohydrate record at the beginning t of the window. In order to minimize the effect of other events, any windows with an insulin or carbohydrate event in time interval $[t-4\text{hours}, t]$ and $[t-2\text{hours}, t]$ respectively will not be considered. The 4-hour filter was chosen for rapid insulin because the effect of Fiasp insulin lasts 3 to 5 hours [23]. For carbohydrates, a 2-hour filter was chosen as this is the time it should take for carbohydrates to digest and turn into blood glucose [24]. Histograms of BG change over 30 minutes and 1 hour following a carbohydrate or insulin event are available in Figure 3.4.

The post-carbohydrate BG change distributions are shifted towards positive values, indicating there is an increase in BG, as expected after consuming carbohydrates. Also, the 1-hour window is shifted more towards positive values than the 30-minute window, therefore BG keeps on increasing after 30 minutes. The same effect is seen in the post-insulin BG change distributions, except it is shifted towards negative values.

One downside of this analysis is that there are not many events where the effects of factors influencing BG levels can be analyzed in isolation. Specifically, it is 47 and 38 events for carbohydrates for 30-minute and 1-hour windows respectively, and 60 and 33 events for rapid insulin.

3.2.1 Blood Glucose Change Behavior

Instead of looking only at the BG change between two time instances, in Figure 3.5 and Figure 3.6 complete curves of BG behavior are shown. The logic for filtering remains the same as previously for the BG change distributions, but the BG change is plotted over the whole 1-hour window. All windows were moved so that they start at 0. The value plotted is thus the BG change.

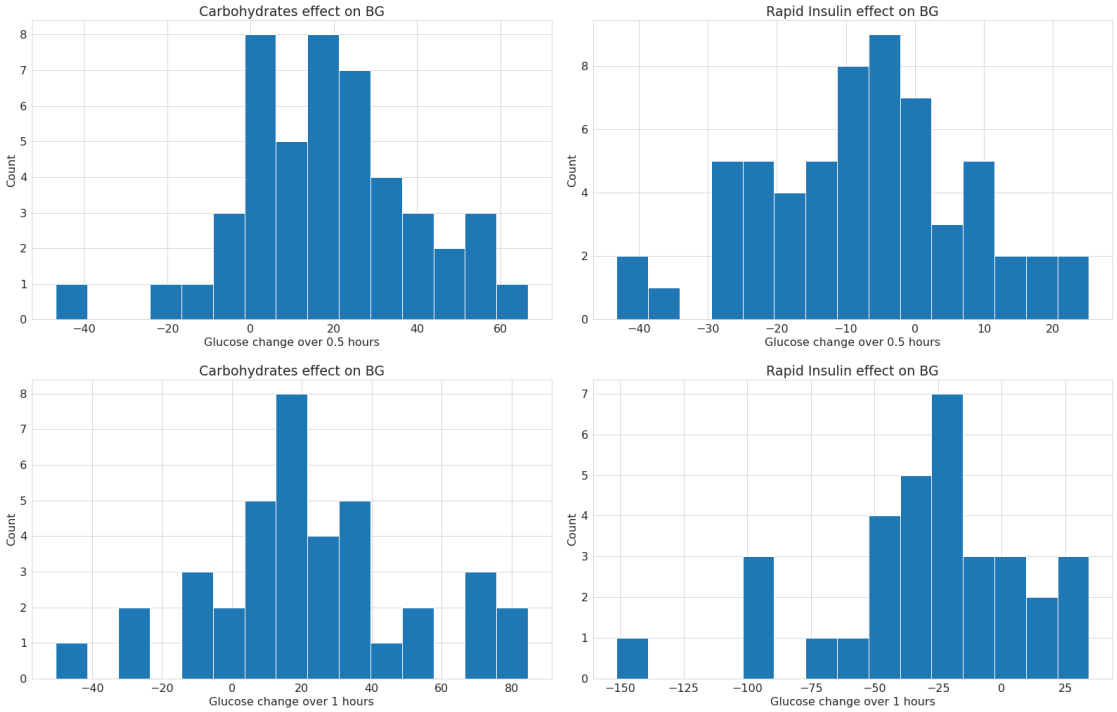

Figure 3.4 Effect of carbohydrates and rapid insulin on BG change

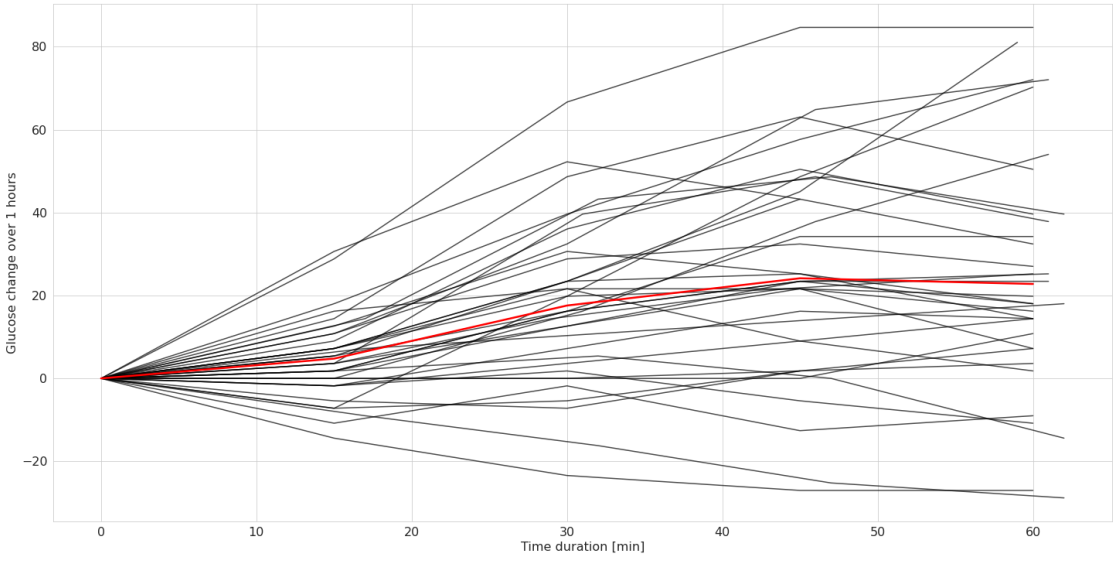

Figure 3.5 BG behavior after carbohydrates consumption with the red-highlighted average

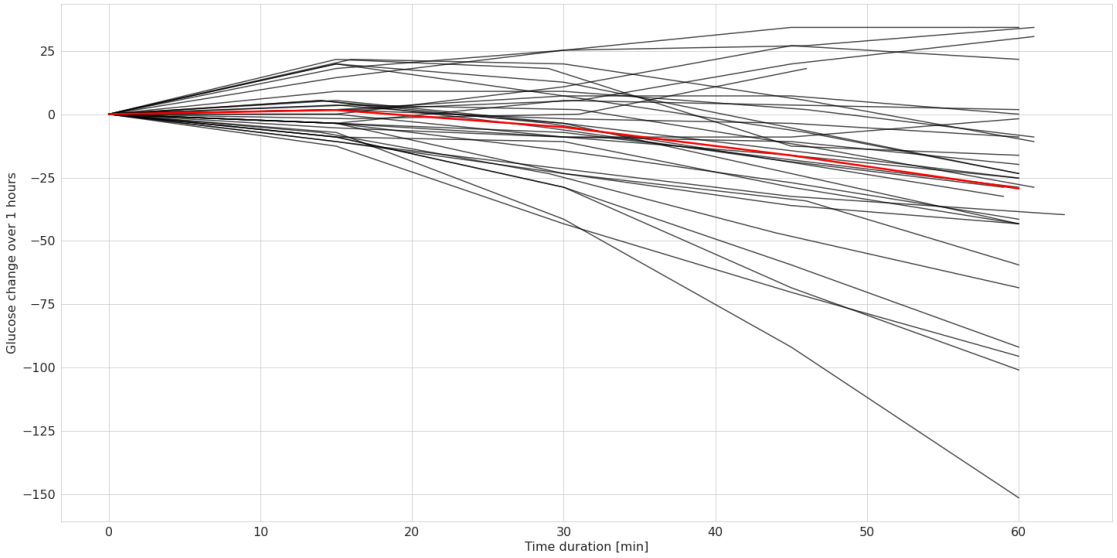

Figure 3.6 BG behavior after rapid insulin injection with red-highlighted average

3.3 The Effect of Physical Activity on BG

Another factor that is important for BG dynamics is physical activity. The most readily available physical activity indicators on fitness trackers are distance, heart rate, and on more advanced trackers, energy expenditure. Investigation of the impact of these features on BG levels is assessed.

Table 3.2 shows basic descriptive statistics for these features, all of which were resampled to 15-minute frequency.

Table 3.2 Descriptive statistics of Distance, Heart rate, and Calories Burned

	Calories [kcal]	Distance [meters]	Heart rate [bpm]
count	12289	12289	12289
mean	26,49	65,37	71,07
std	15,76	148,62	13,24
min	6,46	0	47,02
max	135,29	1151,1	142,87
25%	16,4	0	60,31
50%	19,94	4,8	69,53
75%	29,61	54,8	78,84

To see if there is a relationship between physical activity and BG levels, the correlation between physical activity and BG levels is examined. Specifically, it is the correlation between current BG and 1-hour and 3-hour physical activity history. The sum is used as an aggregating function for distance and calories burned and mean for heart rate. The correlations can be seen in Figure 3.7.

There is not a strong correlation between BG levels and previous physical activity. it was mentioned in Chapter 1 that physical activity may both increase and decrease BG, depending on the nature of the activity. It is possible that this is the reason for the weak correlation observed. However, there is a strong correlation between distance, calories, and bpm.

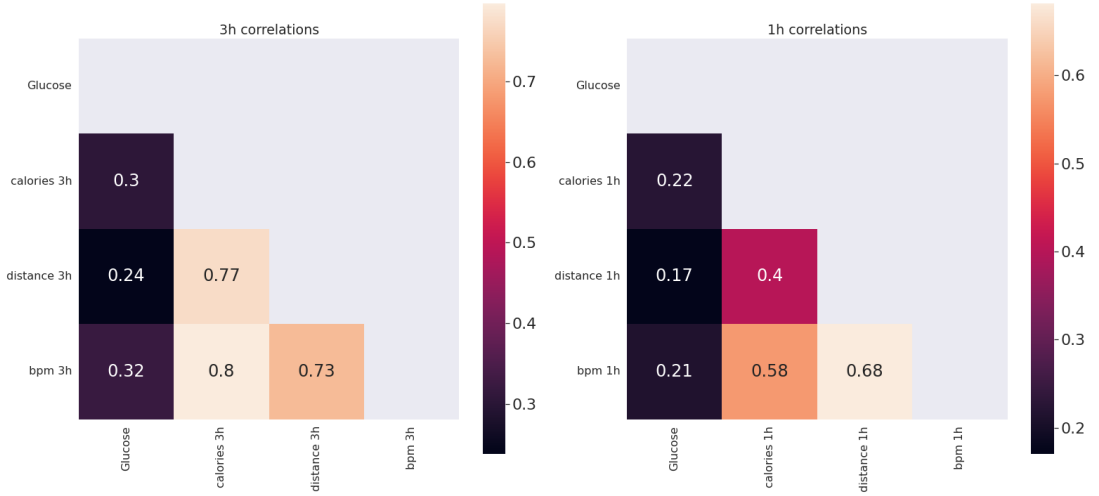

Figure 3.7 Correlation between BG levels and previous physical activity

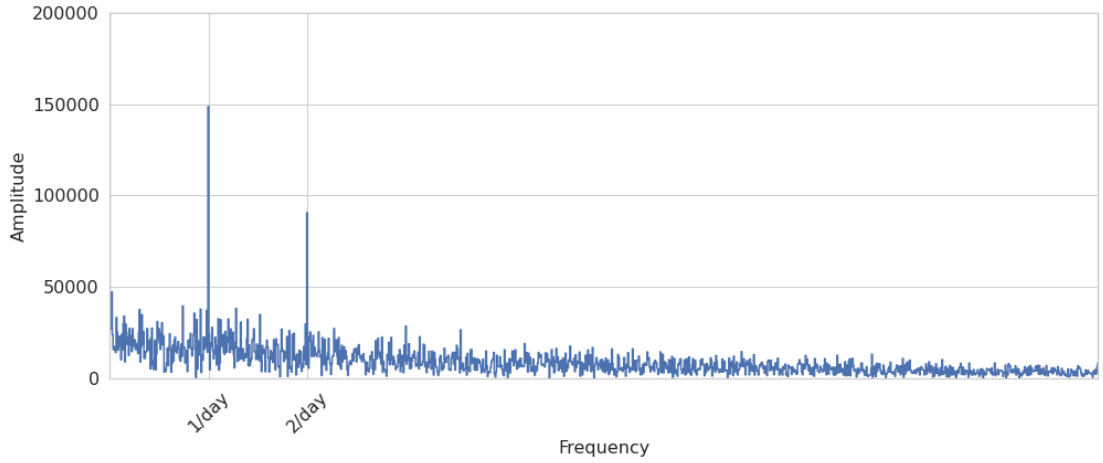

Figure 3.8 BG Fast Fourier transform

3.4 Blood Glucose Periodicity

Meals, insulin injections, and times of higher physical activity tend to have daily patterns. Furthermore, E. Van Cauter et al. [25] have shown that circadian rhythm and sleep influence BG levels as well. They further discovered that periodicity may also come from hormones like the growth hormone, whose increased morning secretion was correlated with an absolute rise in BG levels.

BG periodicity is assessed in Figure 3.8 by calculating a Fast Fourier transform over BG levels sampled in a 15-minute frequency. The amplitudes are significantly higher at 1-day and half-day frequencies suggesting daily and day-night periodicity.

Chapter 4

Model Implementation & Evaluation

ANN-based models and SVMs were implemented and evaluated. In particular, it is feed-forward neural networks (FFNN), recurrent neural networks (RNN), and SVM regression models with radial basis function (RBF) kernel. Both ANNs and SVMs with RBF kernel have the ability to model non-linear phenomena like BG dynamics. Implementation, training, and evaluation of ANN models was realized with the support of the Tensorflow library [26] and with the Scikit-learn library [27] for SVM-based models. The evaluated prediction horizons are 30 minutes, 1 hour, and 2 hours.

4.1 Model Evaluation Strategy

Two metrics are used to quantify the prediction performance of models. A Root mean squared error (RMSE) and for clinical accuracy, it is the Clarke error grid analysis (CEGA) [28].

For a series of BG measurements Y and BG predictions $\hat{Y}$, both having an equal length n, RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \tag{4.1}$$

CEGA is a common tool used to evaluate the clinical accuracy of BG prediction models. It is a scatterplot of reference BG values and predicted BG values and is divided into five zones.

- Zone A contains predicted BG values that are no more than 20% deviated from the actual BG measurements,
- Zone B contains those points that are outside of 20% but would not lead to inappropriate treatment,
- Zone C are points that would lead to unnecessary treatment,
- Zone D contains points indicating a potentially dangerous failure to detect hypoglycemia or hyperglycemia, and
- Zone E contains clinical errors where hyperglycemia is confused for hypoglycemia and vice versa.

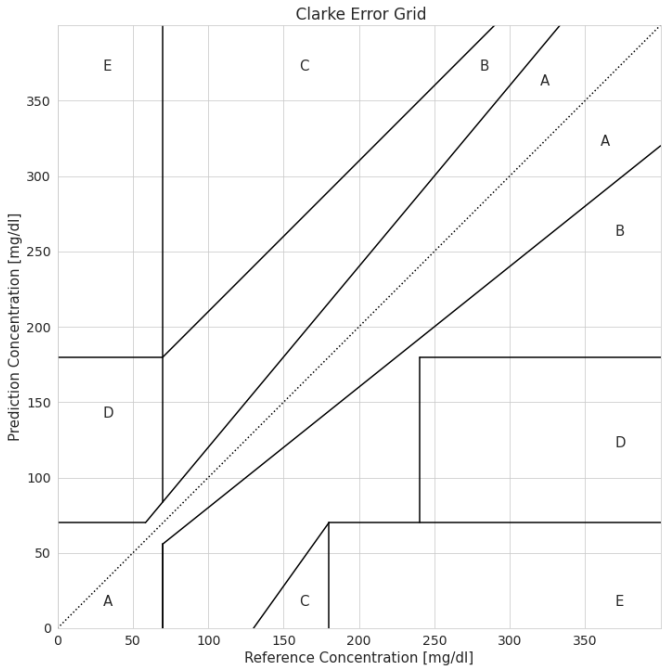

Figure 4.1 Clarke error grid

In short, A is the ideal zone and our target is to have all points in it, B contains benign errors and the rest of the zones should contain as few points as possible. An example of how CEGA looks can be seen in Figure 4.1.

4.2 Data Manipulation

For the purpose of this section, the following capitalized terms are used when talking about dataset features: RapidInsulin, LongInsulin, Carbohydrates, Distance, Calories – calories burned, Bpm, and BG.

4.2.1 Preprocessing

Firstly, individual data sources (the CGM readings, data from an electronic diary, and physical activity from a fitness tracker) are concatenated into one. Afterward, all features not utilized by the models are thrown away.

The next step is to identify missing values. As mentioned in Chapter 2, the glucose sensor needs to be replaced every 14 days. The newly inserted sensor needs a 1-hour warm-up period before it can start providing BG readings. Additionally, if the sensor is not read with the reader at least every 8 hours, data gets overwritten and lost. This may happen if a patient sleeps for more than 8 hours. In such a case, some BG readings may be missing at the sleep onset. Glucose readings are expected every 15 minutes but occasionally may get delayed. Any gaps between BG measurements larger than 20 minutes (1 third of the expected sampling rate) are identified. The total amount of such gaps was 61, with the average gap duration being 1 hour and 14 minutes. Time periods without BG readings make up approximately 2,5 % of the whole 128-day-wide dataset.

Gaps in physical activity were identified by first up-sampling to a 15-minute sampling rate (the rate of CGM) and then looking at samples with no data. There were no missing data for

Calories and Distance and only 0,7 % of missing Bpm readings. All missing data is filled using linear interpolation, but for BG values, interpolation is done only on the train and validation dataset, not on the dataset used for final testing.

4.2.2 Feature Extraction

In Chapter 3 daily BG periodicity was identified, for that reason, the hour of day feature is added with values from 1 to 24, it will be referred to just as Hour.

As stated in Chapter 2 since 9. 1. 2022 the patient recorded values of GI associated with the consumed meal. GI notes are first transformed to numeric values using the following mapping: $Low \rightarrow 15, Medium \rightarrow 45, High \rightarrow 75, VeryHigh \rightarrow 95$. GI values are then transformed into a new feature GlycemicLoad, calculated as

$$\text{GlycemicLoad}_i = \frac{(\text{GI}_i \cdot \text{Carbohydrates}_i)}{100} \tag{4.2}$$

for any meal i recorded in the electronic diary.

All features are resampled to a 15-minute sampling rate. The sum is used as an aggregating function for RapidInsulin, LongInsulin, Carbohydrates, GlycemicLoad, Distance, and Calories. For Bpm, Hour, and BG, the mean aggregating function is used. Thus, an evenly spaced multivariate time series is created, which will be easier to work with later on.

To capture data about insulin sensitivity, a RapidInsulin6d feature is created, which contains the 6-day mean total daily insulin dose, and any data missing at the first 6 days of the dataset is back-filled.

As stated in Chapter 1, the long insulin used in this study has an approximately 24-hours long activity period. So that it is known at any point in time, what was the last LongInsulin dose, the last LongInsulin dose is forward-filled over the dataset until the next LongInsulin dose, and so on all the way to the end of the dataset.

4.2.3 Normalization and Split

The data is split into a training, validation, and testing dataset, and each feature value X_i at time i is min-max normalized.

$$X_i = \frac{X_i - X_{\min}}{X_{\max} - X_{min}} \tag{4.3}$$

where $X_{\min}$ and $X_{\max}$ are respectively, the minimum and the maximum value of feature X seen in the training dataset.

The train-validation-test split percentages are 70 %, 20 % and 10 %. There are 8602, 2458, and 1187 rows in the training, validation, and testing dataset respectively, which translates to 89, 25, and 12 complete days. Data is not being randomly shuffled before the split. It ensures that slicing the data into windows of consecutive samples will be possible and it makes validation and test results more realistic, as models are evaluated on data collected after the model training.

4.2.4 Insulin Activity Curves

Georga et al. [10] used meal and insulin compartmental models instead of raw carbohydrate and insulin samples. The meal compartmental model estimates the rate of appearance of meal-derived glucose in the systemic circulation [29] and the insulin compartmental model estimates the plasma insulin concentration [30]. They used values from these models that expand up to the time for which the prediction is to be made, i.e., $t + PH$. The upcoming values within the time interval $[t, t + PH]$ are computed by the compartmental models using the insulin and meal recordings until the current time t [10].

Due to the complexity of these mathematical compartmental models and difficulty finding appropriate parameter values for them, we decided to use a simplistic insulin model, inspired by the DIY closed-loop system LoopKit [31]. The model used is capable of estimating the active insulin in the body by using an exponential decay function that models the insulin activity. This function will be different between different types of insulin and is determined by two parameters, td and tp. td is the total duration in minutes when the injected insulin is active and tp is the time to peak insulin activity in minutes. The IOB(t) (insulin-on-board) function is defined the same as in LoopKit [31]. It takes a parameter t, the time since insulin injection, and returns a value from interval $[0, 1]$ which is the estimated percentage of remaining active insulin.

$$\begin{aligned} \tau &= tp\frac{(1-\frac{tp}{td})}{1-2\frac{tp}{td}} \\ a &= 2\frac{\tau}{td} \\ s &= \frac{1}{1-a+(1+a)\cdot\exp(-\frac{td}{\tau})} \\ \mathrm{IOB}(t) &= 1-s(1-a)\Big(\Big(\frac{t^2}{\tau\cdot td\cdot(1-a)}-\frac{t}{\tau}-1\Big)\cdot\exp\Big(-\frac{t}{\tau}\Big)\Big) \end{aligned} \tag{4.4}$$

where τ is the time constant of exponential decay, a is the rise time factor and s is the auxiliary scale factor.

IA(t) is a function for insulin absorption and shares the td and tp parameters with IOB(t). td and tp should be chosen so that IA(t) models the real insulin absorption as close as possible.

$$\mathrm{IA}(t) = \frac{s}{\tau^2}\cdot t\cdot\Big(1-\frac{t}{td}\Big)\cdot\exp\Big(-\frac{t}{\tau}\Big) \tag{4.5}$$

Figure 4.2 shows plotted IOB(t) and IA(t) with parameters $ta = 55$ and $td = 300$ which were picked for our insulin model.

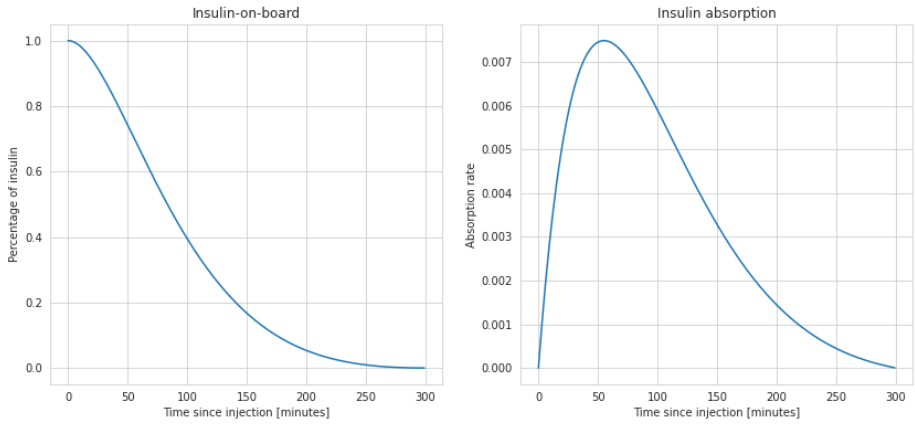

Figure 4.2 Insulin-on-board and insulin absorption curves

A new feature called rapid insulin IOB (RIOB) is created and is calculated as follows. First a new column is created for the feature with the same length as the entire dataset and is filled with zeros. Next, for each non-zero RapidInsulin sample D at time t, a slice of RIOB values is taken from time interval $[t, t + td]$ and each RIOB_i for $i \in [0, \mathrm{td}]$ from the slice is filled as $\mathrm{RIOB}_i = \mathrm{RIOB}_i + D \cdot \mathrm{IOB}(i)$. As visible in the equation, if RIOB_i is not zero, we do not

overwrite, but add to it, so that if insulin activity from two or more separate insuline injections overlap, we compound the expected RIOB. The quantity measured by RIOB is the estimate of insulin units active in the circulation.

The benefit of RIOB is that our models can now use "future values" as inputs. If a prediction is to be made at time t of BG at time $t + PH$, RIOB can be calculated for the whole $[t, t + PH]$ time interval from rapid insulin samples seen before t. The next benefit is that the insulin model is defined by only 2 values, which makes it much simpler than compartmental models.

4.2.5 Creating windows

For any time t and PH, our models should predict BG value at time $t + PH$ based on historical samples seen before time t. In order to do that, there must be a way of preparing windows of historical values as model inputs and future BG as model targets. For that reason, a WindowGenerator class was created. It can create windows of input and output tensors in the form of a TensorFlow dataset. The main parameters of WindowGenerator are:

- The features used (BG, Carbohydrates, etc.).
- Width of the input (in time steps), could also a be a list of widths, with each width corresponding to a different feature.
- The PH, also measured in number of time steps, e.g. PH of 4 means 1 hour (15-minute sampling frequency),
- Batch size – ANN-based models are usually trained on batched data, this parameter sets the number of windows per batch.

The windows are created with a stride equal to 1. So, for a total window size equal to 5, the window intervals w_i will be $w_0 = [0, 4]$, $w_1 = [1, 5]$, ...

If all features have an equal input width, the shape of the input tensor is $(\text{BatchSize}, \text{InputWidth}, \text{NumberOfFeatures})$, where the BatchSize refers to the number of windows in a single batch, InputWidth is the number of historical samples used and NumberOfFeatures is how many dimensions are used to represent data in one time step (sample). However, if the input width is not the same for all features the shape will be

$$\left(\text{BatchSize}, \sum_{i=1}^{n} \text{InputWidth}_i\right) \tag{4.6}$$

for n being the number of features used. In other words, for variable feature input widths, individual feature inputs are flattened and then concatenated into a single tensor.

Another feature of the WindowGenerator class is that it filters out windows with noise in the time interval $[t, t + PH]$. Noise is anything that can significantly affect the BG levels and therefore would alter the mapping of inputs seen until time t onto the BG at time $t + PH$ as an output. The noise events are insulin injection, carbohydrate intake, and higher physical activity. The threshold for high physical activity is a calory output of 81 kcal per 15 minutes. Approximately 1,8 % of our samples are above this threshold.

WindowGenerator also filters out any windows containing gaps of missing BG data, this is only necessary for the testing dataset, as it is not interpolated.

Without dropping noisy windows and windows with missing data there are approximately 8600, 2460, and 1190 windows for training, validation, and testing respectively. This drops down by approximately 64 %, 45 %, and 29 % for 2-hours, 1-hour, and 30-minutes PHs respectively after the reduction.

4.2.6 Feature Selection

The quality of some of the features is assessed before creating and optimizing hyper-parameters of the models. This would ideally be done as part of the model optimization process, where for each model type, the best-performing set of features and input widths would be identified. However, it would significantly increase the required computational resources. For that reason, the quality of the features is assessed either in pairs or by comparing the performance of a model before and after adding a new feature.

The quality of the RIOB feature was assessed in pair with the RapidInsulin feature. Two LSTM models were used, both having the same hyper-parameters and the same window input width. An approximate 3 % gain in prediction performance was seen when using RIOB compared to the RapidInsulin feature.

Pair assessment was also performed on GlycemicLoad and Carbohydrate features using FFNN models, both having the same parameters and input widths. There was no performance gain in using GlycemicLoad. The failure of GlycemicLoad to increase performance might be attributed to low quality of GI notes or to the fact that GI was collected only after 9. 1. 2022.

A reference FFNN model trained on windows with BG, Carbohydrates, and RIOB as features was compared with 3 other models where RapidInsulin6d, LongInsulin, and Hour were added in isolation, all having an input width of 1. There was about 1 % performance gain when using LongInsulin and Hour, but no performance was gained when using RapidInsulin6d.

Lastly, the same reference FFNN model was compared to a model where calories were added as input. There were no performance gains seen when using the calories feature, even when the reference model parameters were altered by adding more neurons to the hidden layer. The failure of calories to increase the prediction performance was seen again in LSTM models. 2 LSTM models were created, both having the same window input width, and with only one having access to the calories feature. Hyper-parameter tuning was performed on both models and no performance gain was seen in the LSTM model using calories.

4.3 Model Selection and Evaluation

For each of the ML models implemented, hyper-parameters tuning is performed. ANN-based models are tuned using a relatively novel HyperBand hyper-parameter tuner [32] implemented in TensorFlow. Each ANN model has a maximum of 70 training epochs with early stopping if validation loss fails to decrease in 2 consecutive epochs. The early stopping helps us save resources as well as keep the model from overfitting the training data. HyperBand algorithm is always run 8 times. All models use adaptive learning rate optimizer Adam [33]. The learning rate for Adam is determined experimentally using the HyperBand tuner with $1 \cdot 10^{-2}, 1 \cdot 10^{-3}, 1 \cdot 10^{-4}$, being the possible values.

4.3.1 Baseline

Before implementing any more complicated models for BG prediction, a naive baseline is constructed as a point of comparison. The baseline model assumes BG will remain the same and it always returns the current BG as a prediction. It takes time for BG to change, so for short PHs, this assumption is not a bad starting point. Of course, for longer PHs the baseline will not work very well. The baseline was evaluated on the test dataset. The results for 30-minute, 1-hour, and 2-hour PHs can be seen in Table 4.1.

Table 4.1 Baseline performance evaluation by PH

PH	RMSE [mg/dl]	Zone A	Zone B	Zone C	Zone D	Zone E
30 minutes	22,99	0,83	0,16	0	0,01	0
1 hour	33,12	0,66	0,30	0,005	0,03	0,005
2 hours	43,69	0,44	0,42	0,02	0,12	0

4.3.2 Feed Forward Neural Networks

The proposed FFNN uses Rectified Linear Unit (ReLU) activation function in its hidden layers and a linear activation function in the output layer with one neuron. The number of hidden layers as well as the number of neurons in each layer is determined in hyper-parameter tuning. Another hyper-parameter is dropout after each hidden layer determined by the dropout rate, which can be either 0; 0,05; 0,10, or 0,15. Dropout is used in order to prevent the model from overfitting.

For each PH, the FFNN models were trained on the following features: BG, Carbohydrates, LongInsulin, Hour, and RIOB with feature input widths: 8 (2 hours) for BG and Carbohydrates, 1 for Hour and LongInsulin, and 8 (2 hours) of estimated future RIOB values. Individual feature inputs are flattened and concatenated into a single tensor input.

The optimal hyper-parameters for 2-hours, 1-hour, and 30-minutes PHs can be seen in Table 4.2.

Table 4.2 Optimal FFNN hyper-parameters per each PH

PH	1. layer neurons	2. layer neurons	1. dropout	2. dropout	learning rate
30 minutes	40	0	0	0	$1 \cdot 10^{-3}$
1 hour	48	8	0	0	$1 \cdot 10^{-3}$
2 hours	40	24	0,05	0,1	$1 \cdot 10^{-3}$

The performance of the best FFNN model for each PH can be seen in Table 4.3.

Table 4.3 FFNN performance evaluation by PH

PH	RMSE [mg/dl]	Zone A	Zone B	Zone C	Zone D	Zone E
30 minutes	18,26	0,87	0,12	0	0,01	0
1 hour	25,52	0,73	0,21	0	0,06	0
2 hours	27,11	0,52	0,27	0	0,21	0

4.3.3 Recurrent Neural Networks

A Long-short-term memory (LSTM) recurrent neural network is used. It was developed by Sepp Hochreiter and Jürgen Schmidhuber [34]. The property of LSTM is that it can store information over extended time intervals, which makes it suitable for applications in machine translation, image captioning, and time-series forecasting.

In contrast with the FFNN, inputs can not be flattened before passing them as an input to the network. The LSTM network expects the input to be a 3-dimensional tensor with a shape (BatchSize, InputWidth, NumberOfFeatures). Because of this limitation, LongInsulin and Hour features were not applicable. So, for each PH, the LSTM models were trained on BG, Carbohydrates, and RIOB. The input width is 20 (5 hours) for all of the features.

The main LSTM parameter is ”units”, this however does not refer to the number of LSTM cells nor the number of time steps, but to the dimension of state output from the LSTM cell.

Figure 4.3 shows a graphical interpretation. Units parameter is chosen in hyper-parameter tuning. The dropout rate is determined in hyper-parameter tuning as well and has the same domain of values as FFNN models.

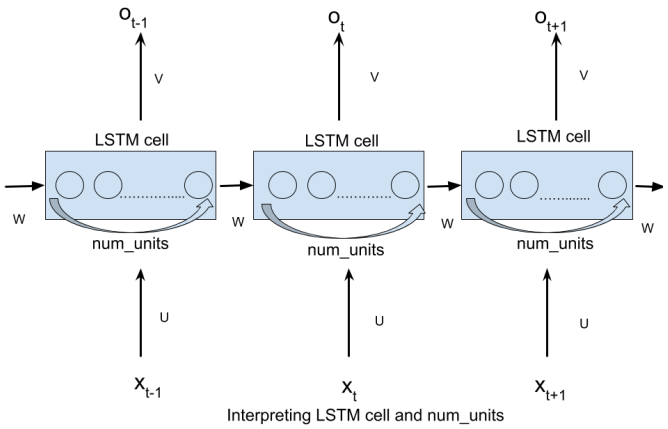

Figure 4.3 LSTM architecture and explanation of number of units parameter

The optimal LSTM hyper-parameter can be seen in Table 4.4.

Table 4.4 Optimal LSTM hyper-parameters per each PH

PH	units	dropout rate	learning rate
30 minutes	32	0,15	$1 \cdot 10^{-3}$
1 hour	32	0	$1 \cdot 10^{-2}$
2 hours	56	0,05	$1 \cdot 10^{-2}$

The performance of the best LSTM model for each PH can be seen in Table 4.5. Furthermore, visualizations of predictions for 2-hours PH are shown in Figure 4.4. Figure 4.5 shows multiple consecutive predictions made with 30-minute PH on a slice of the time series.

Table 4.5 LSTM performance evaluation by PH

PH	RMSE [mg/dl]	Zone A	Zone B	Zone C	Zone D	Zone E
30 minutes	17,06	0,9	0,09	0	0,01	0
1 hour	24,32	0,78	0,19	0	0,03	0
2 hours	30,2	0,55	0,27	0	0,18	0

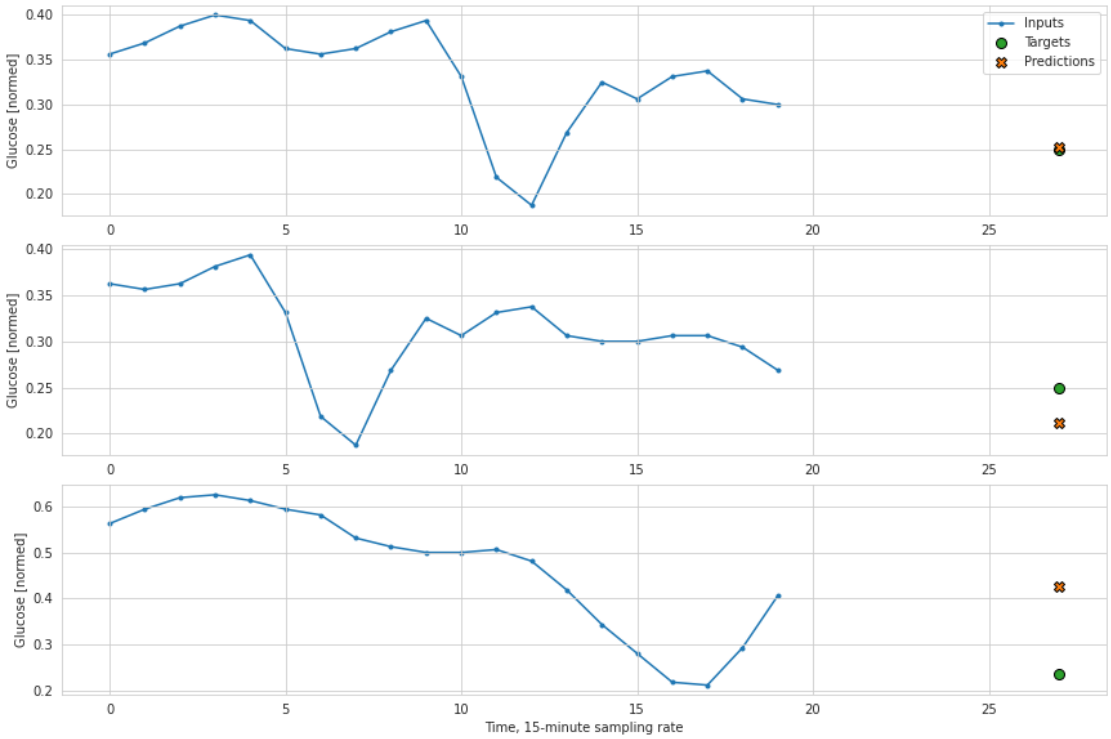

Figure 4.4 LSTM predictions for 2-hours PH

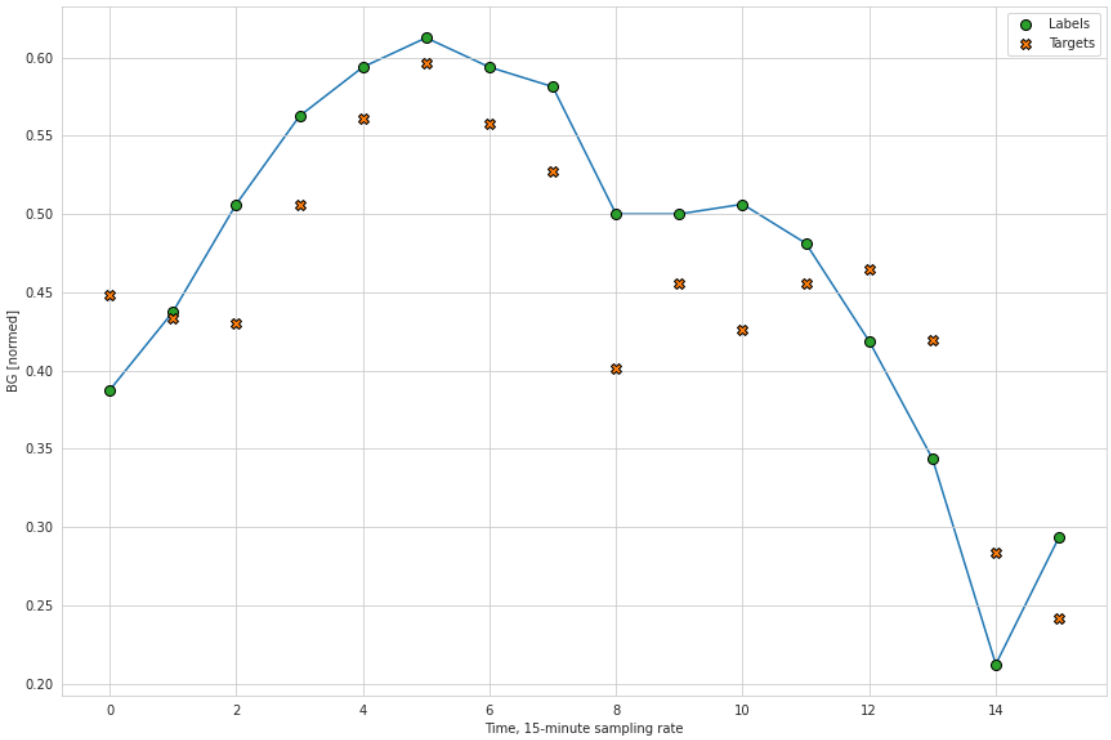

Figure 4.5 Multiple LSTM predictions for 30-minute PH

4.3.4 Support Vector Machines

SVMs are investigated as an alternative to ANN-based approaches. In particular, it is support vector regression [35], as the issue of BG prognosis is estimating a real number (the BG level at some point in time). SVMs may be beneficial in cases of limited datasets and are computationally less expensive than the ANN-based approaches. The same as ANNs, SVMs are capable of modeling non-linear phenomena, like BG dynamics.

RBF kernel function is used in the SVM and the parameters are C, ϵ and γ. ϵ is the maximal error, where no penalty is associated in the training loss function with points predicted within a distance ϵ from the actual value. C is the regularization parameter, and simply speaking, as C increases, our tolerance for points outside of ϵ also increases. Lastly, γ is a RBF kernel coefficient. The inputs of SVM models are exactly the same as inputs of FFNN models.

SVM hyper-parameters were tuned by simply iterating over all combinations and choosing the best-performing one. The optimal SVM hyper-parameters can be seen in Table 4.6.

Table 4.6 Optimal SVM hyper-parameters per each PH

PH	γ	ϵ	C
30 minutes	0,27	0,02	1
1 hour	1	0,02	1
2 hours	1	0,06	1

The performance of the best SVM model for each PH can be seen in Table 4.7.

Table 4.7 SVM regression performance evaluation by PH

PH	RMSE [mg/dl]	Zone A	Zone B	Zone C	Zone D	Zone E
30 minutes	17,41	0,9	0,09	0	0,01	0
1 hour	24,9	0,8	0,17	0	0,03	0
2 hours	27,48	0,6	0,25	0	0,15	0

4.3.5 Multi-output Predictions

The goal of the thesis was to predict single BG values at a pre-defined PH, however, it is possible to predict not only the single BG value but also a BG curve – multiple consecutive future BG values. Proposed ANN-based models can be simply modified to predict BG curves, by adding more output neurons, one for each future BG value. Figure 4.6 shows a plot of targets and predictions from a LSTM model trained to predict the next 8 BG values (2 hours).

4.4 Results

RMSE results indicate that the best model for a 30-minute PH is LSTM with 17,06 mg/dl RMSE. LSTM also performs the best on a 1-hour PH with 24,32 mg/dl RMSE. However, for 2-hours PH, the best model is a FFNN with 27,11 mg/dl RMSE. When inspecting which models had the highest percentage of points in Zone A of the CEGA grid, SVMs had the best performance for all PHs, having 90 %, 80%, and 60 % of points in Zone A respectively for 30-minute, 1-hour, and 2-hours PHs. The individual CEGA plots for each PH can be seen in Figure 4.7, 4.8, and 4.9.

When comparing the performance achieved with our models to the performance of existing approaches, our models tend to perform better in the long-term PHs and worse in the short-term 30-minute PH. The only publications with a better result for 2-hours PH are the ones from

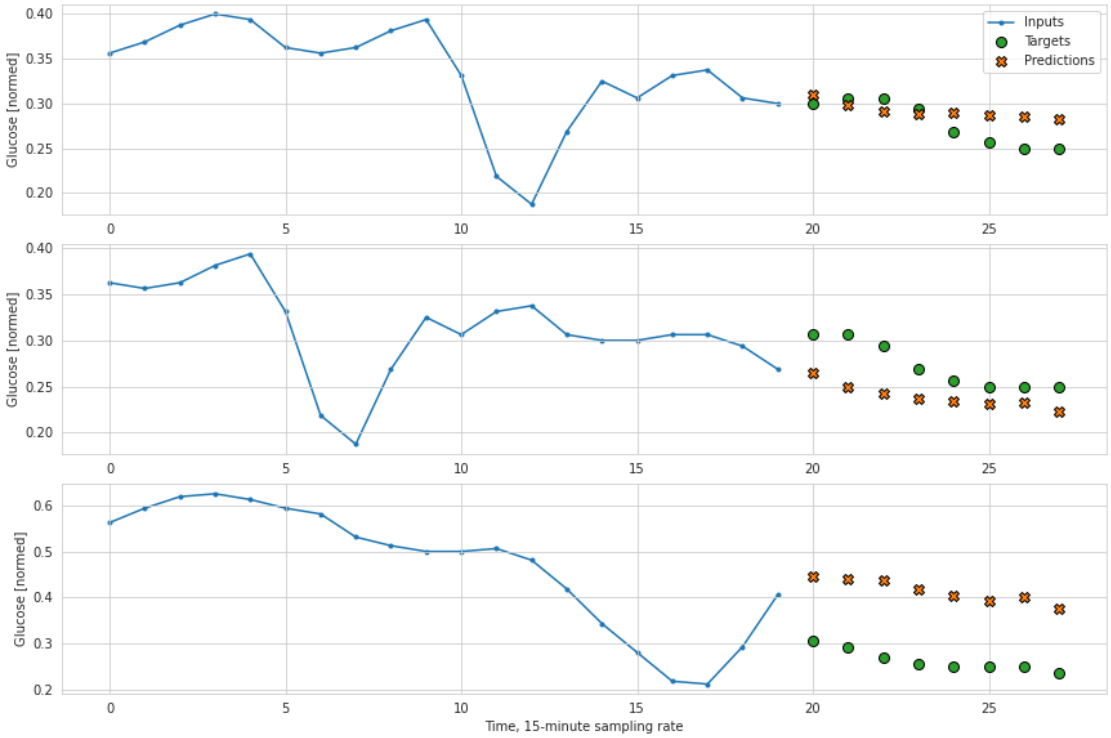

Figure 4.6 Multi-output prediction of next 8 BG samples

Georga et al. [10] [36]. Our results for 2-hours PH outperform all of the other publications from the review by Ashenafi Zebene Woldaregay et al. [7] as seen in Figure 1.2.

Usage of insulin and meal models capable of modeling the insulin activity and estimating the velocity of meal-derived glucose release seems to help with the prediction performance. We have also verified this by seeing better results with the RIOB feature than with raw RapidInsulin samples.

We believe that further improvement to the performance can be made by using a more granular sampling frequency. All of the used features were resampled to a 15-minute sampling frequency as this is the sampling frequency of the CGM sensor. This causes information loss, especially in RapidInsulin and Carbohydrates features, because the time of insulin injection or meal consumption will be less precise.

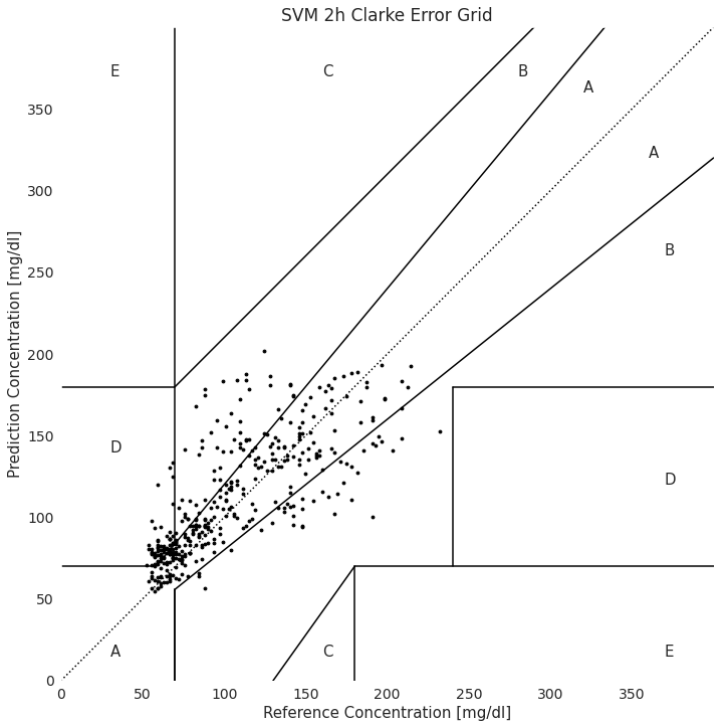

Figure 4.7 SVM CEGA plot for 2-hour PH

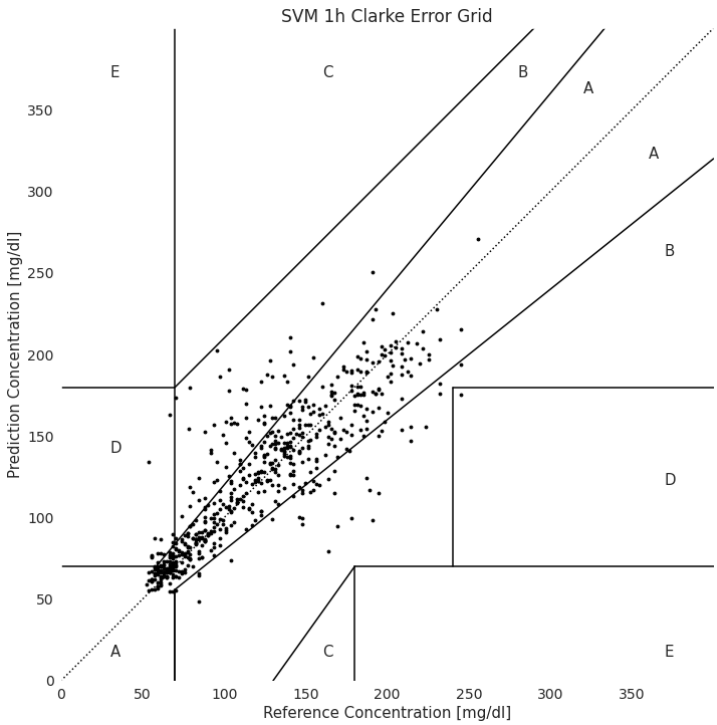

Figure 4.8 SVM CEGA plot for 1-hour PH

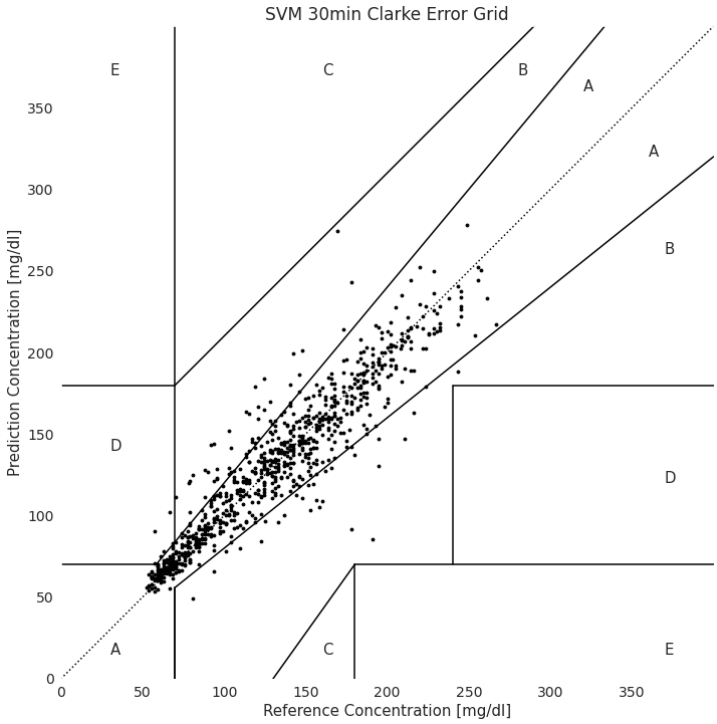

Figure 4.9 SVM CEGA plot for 30-minute PH

Chapter 5
Conclusion

The goal of this thesis was to analyze BG dynamics, research and evaluate models for BG prediction and to collect a dataset containing BG levels, insulin and carbohydrates intake notes, and physical activity.

The author fully met all of these goals. Firstly a rich and extensive dataset with features relevant to BG analysis and prediction was collected. The dataset contains over 4 months' worth of data collected in free-living conditions. We went further than required by adding additional features to the dataset.

We analyzed the dataset in order to identify and quantify the effects of various factors on BG levels. It was done by studying how BG changes in response to these factors in isolation.

We implemented ANN-based and SVM-based models for BG prediction for 30-minute, 1-hour, and 2-hours prediction horizons. With regard to the reviewed academic literature on BG prediction, our models for 2-hours PH are amongst the best-performing.

To our knowledge, exponential decay functions as an insulin model were not previously used with SVMs and ANNs. We explored them as an alternative to compartmental models and confirmed that they can improve the performance of ML models.

Since our prediction models are solely trained on data from consumer-grade hardware captured in free-living conditions, we believe that there is a clear path forward to start using them as a real-time decision-support system. Of course, it must be done with due diligence and caution as models still make errors, but the clinical accuracy, especially for short PHs, suggests these errors are in a tolerable range. We believe, that the prediction performance can be further improved with new approaches and by using more and/or different features. That is why we decided to make our dataset available publicly to foster open collaboration and competition in developing the best models for BG prediction.

Bibliography

1. *Glucose Tolerance test* [online] [visited on 2022-04-18]. Available from: www.medlineplus.gov/ency/article/003466.htm.
2. CRYER, Philip E.; DAVIS, Stephen N.; SHAMOON, Harry. Hypoglycemia in Diabetes. *Diabetes Care*. 2003, vol. 26, no. 6, pp. 1902–1912. ISSN 0149-5992. Available from DOI: 10.2337/diacare.26.6.1902.
3. *Hyperglycemia* [online] [visited on 2022-04-18]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK430900/.
4. FOSTER-POWELL, K; MILLER, J B. International tables of glycemic index. *The American Journal of Clinical Nutrition*. 1995, vol. 62, no. 4, 871S–890S. ISSN 0002-9165. Available from DOI: 10.1093/ajcn/62.4.871S.
5. DAVIDSON, Paul C; HEBBLEWHITE, Harry R; STEED, Robert D; BODE, Bruce W. Analysis of guidelines for basal-bolus insulin dosing: basal insulin, correction factor, and carbohydrate-to-insulin ratio. *Endocrine Practice*. 2008, vol. 14, no. 9, pp. 1095–1101. Available also from: https://doi.org/10.4158/EP.14.9.1095.
6. KROLL, Martin H. Biological variation of glucose and insulin includes a deterministic chaotic component. *Biosystems*. 1999, vol. 50, no. 3, pp. 189–201. ISSN 0303-2647. Available from DOI: https://doi.org/10.1016/S0303-2647(99)00007-6.
7. WOLDAREGAY, Ashenafi Zebene; 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. *Artificial Intelligence in Medicine*. 2019, vol. 98, pp. 109–134. ISSN 0933-3657. Available from DOI: https://doi.org/10.1016/j.artmed.2019.07.007.
8. BREMER, Troy; GOUGH, David A. Is blood glucose predictable from previous values? A solicitation for data. *Diabetes*. 1999, vol. 48, no. 3, pp. 445–451. Available also from: https://diabetesjournals.org/diabetes/article/48/3/445/12150/Is-blood-glucose-predictable-from-previous-values.
9. BEN ALI, Jaouher; HAMDI, Takoua; FNAIECH, Nader; DI COSTANZO, Véronique; FNAIECH, Farhat; GINOUX, Jean-Marc. Continuous blood glucose level prediction of Type 1 Diabetes based on Artificial Neural Network. *Biocybernetics and Biomedical Engineering*. 2018, vol. 38, no. 4, pp. 828–840. ISSN 0208-5216. Available from DOI: https://doi.org/10.1016/j.bbe.2018.06.005.

10. GEORGA, Eleni I.; PROTOPAPPAS, Vasilios C.; ARDIGÒ, Diego; MARINA, Michela; ZAVARONI, Ivana; POLYZOS, Demosthenes; FOTIADIS, Dimitrios I. Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression. *IEEE Journal of Biomedical and Health Informatics*. 2013, vol. 17, no. 1, pp. 71–81. Available from DOI: 10.1109/TITB.2012.2219876.

11. *Kaggle website* [online] [visited on 2022-04-16]. Available from: https://www.kaggle.com/.

12. *Kaggle Type 1 Diabetes dataset* [online] [visited on 2022-05-09]. Available from: www.kaggle.com/datasets/lacofloris/type-1-diabetes-blood-glucose-prediction.

13. ABBOTT DIABETES CARE INC. *FreeStyle LibreLink - DE* [comp. software]. 2021. Version 2.5.3. Available also from: https://freestyle.abbott.

14. BASU, Ananda; DUBE, Simmi; SLAMA, Michael; ERRAZURIZ, Isabel; AMEZCUA, Jose Carlos; KUDVA, Yogish C.; PEYSER, Thomas; CARTER, Rickey E.; COBELLI, Claudio; BASU, Rita. Time Lag of Glucose From Intravascular to Interstitial Compartment in Humans. *Diabetes*. 2013, vol. 62, no. 12, pp. 4083–4087. ISSN 0012-1797. Available from DOI: 10.2337/db13-1132.

15. KOVATCHEV, Boris P; SHIELDS, Devin; BRETON, Marc. Graphical and numerical evaluation of continuous glucose sensing time lag. *Diabetes technology & therapeutics*. 2009, vol. 11, no. 3, pp. 139–143. Available also from: https://www.liebertpub.com/doi/full/10.1089/dia.2008.0044.

16. MYSUGR GMBH. *mySugr* [comp. software]. 2022. Version 3.92.28. Available also from: https://mysugr.com.

17. DINE4FIT, A.S. *Kalorické Tabulky* [comp. software]. 2022. Version 3.7.1. Available also from: https://www.kaloricketabulky.cz/.

18. O'DRISCOLL, Ruairi; TURICCHI, Jake; BEAULIEU, Kristine; SCOTT, Sarah; MATU, Jamie; DEIGHTON, Kevin; FINLAYSON, Graham; STUBBS, James. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *British Journal of Sports Medicine*. 2020, vol. 54, no. 6, pp. 332–340. Available from eprint: https://bjsm.bmj.com/content/bjsports/54/6/332.full.pdf.

19. *JupyterLab* [online] [visited on 2022-05-05]. Available from: www.github.com/jupyterlab/jupyterlab.

20. REBACK, Jeff; JBROCKMENDEL; MCKINNEY, Wes; BOSSCHE, Joris Van den; CLOUD, Phillip; HAWKINS, Simon; ROESCHKE, Matthew; GFYOUNG; SINHRKS; KLEIN, Adam; PETERSEN, Terji; AUGSPURGER, Tom; HOEFLER, Patrick; TRATNER, Jeff; SHE, Chang; AYD, William; NAVEH, Shahar; GARCIA, Marc; DARBYSHIRE, JHM; SCHENDEL, Jeremy; HAYDEN, Andy; SHADRACH, Richard; SAXTON, Daniel; GORELLI, Marco Edward; LI, Fangchen; ZEITLIN, Matthew; JANCAUSKAS, Vytautas; MCMASTER, Ali; BATTISTON, Pietro; SEABOLD, Skipper. *pandas-dev/pandas: Pandas 1.3.5* [comp. software]. Zenodo, 2021. Version v1.3.5. Available from DOI: 10.5281/zenodo.5774815.

21. HUNTER, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*. 2007, vol. 9, no. 3, pp. 90–95. Available from DOI: 10.1109/MCSE.2007.55.

22. WASKOM, Michael L. seaborn: statistical data visualization. *Journal of Open Source Software*. 2021, vol. 6, no. 60, p. 3021. Available from DOI: 10.21105/joss.03021.

23. *Fiasp Information* [online] [visited on 2022-05-03]. Available from: https://www.medicines.org.uk/emc/medicine/33022.

24. WOLEVER, Thomas MS. Carbohydrate and the regulation of blood glucose and metabolism. *Nutrition reviews*. 2003, vol. 61, no. suppl_5, S40–S48. Available from DOI: 10.1301/nr.2003.may.S40-S48.

25. VAN CAUTER, Eve; BLACKMAN, John D; ROLAND, Dianne; REFETOFF, Samuel; SPIRE, Jean Paul; POLONSKY, Kenneth S, et al. Modulation of glucose regulation and insulin secretion by circadian rhythmicity and sleep. *The Journal of clinical investigation*. 1991, vol. 88, no. 3, pp. 934–942. Available also from: https://www.jci.org/articles/view/115396.

26. ABADI, Martín; AGARWAL, Ashish; BARHAM, Paul; BREVDO, Eugene; CHEN, Zhifeng; CITRO, Craig; CORRADO, Greg S; DAVIS, Andy; DEAN, Jeffrey; DEVIN, Matthieu, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*. 2016.

27. 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. *Journal of Machine Learning Research*. 2011, vol. 12, pp. 2825–2830.

28. CLARKE, William L; COX, Daniel; GONDER-FREDERICK, Linda A; CARTER, William; POHL, Stephen L. Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose. *Diabetes Care*. 1987, vol. 10, no. 5, pp. 622–628. ISSN 0149-5992. Available from DOI: 10.2337/diacare.10.5.622.

29. LEHMANN, E.D.; DEUTSCH, T. A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. *Journal of Biomedical Engineering*. 1992, vol. 14, no. 3, pp. 235–242. ISSN 0141-5425. Available from DOI: https://doi.org/10.1016/0141-5425(92)90058-S. Annual Scientific Meeting.

30. TARIN, C.; TEUFEL, E.; PICO, J.; BONDIA, J.; PFLEIDERER, H.-J. Comprehensive pharmacokinetic model of insulin Glargine and other insulin formulations. *IEEE Transactions on Biomedical Engineering*. 2005, vol. 52, no. 12, pp. 1994–2005. Available from DOI: 10.1109/TBME.2005.857681.

31. *Loopkit prediction algorithm* [online] [visited on 2022-05-05]. Available from: https://loopkit.github.io/loopdocs/operation/algorithm/prediction/#predicting-glucose.

32. LI, Lisha; JAMIESON, Kevin; DESALVO, Giulia; ROSTAMIZADEH, Afshin; TALWALKAR, Ameet. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *Journal of Machine Learning Research*. 2018, vol. 18, no. 185, pp. 1–52. Available also from: http://jmlr.org/papers/v18/16-558.html.

33. KINGMA, Diederik P; BA, Jimmy. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 2014. Available also from: https://arxiv.org/abs/1412.6980.

34. HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. *Neural Computation*. 1997, vol. 9, no. 8, pp. 1735–1780. Available from DOI: 10.1162/neco.1997.9.8.1735.

35. SMOLA, Alex J; SCHÖLKOPF, Bernhard. A tutorial on support vector regression. *Statistics and computing*. 2004, vol. 14, no. 3, pp. 199–222.

36. GEORGA, Eleni I.; PROTOPAPPAS, Vasilios C.; POLYZOS, Demosthenes; FOTIADIS, Dimitrios I. A predictive model of subcutaneous glucose concentration in type 1 diabetes based on Random Forests. In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2012, pp. 2889–2892. Available from DOI: 10.1109/EMBC.2012.6346567.

Contents of the attached media

- `figures` ..... all of the figures and visualisations generated
- `models` ..... exports of all best-performing trained ML models
- `dataset.py` ..... python source code for reading the data exports
- `fitbit_visualisations.ipynb` ..... visualisations of Fitbit data
- `visualisations.ipynb` ..... visualisations used in Chapter Analysis
- `prediction_models.ipynb` ..... data preprocessing and ML model creation
- `requiremets.txt` ..... list of dependencies
- `data`
  - `fitbit_data` ..... exports from Fitbit fitness tracker
  - `mySugr_data` ..... exports from mySugr mobile app
  - `reader_data` ..... exports from CGM Freestyle libre
- `thesis_text`
  - `thesis.pdf` ..... text of the thesis in PDF format
  - `thesis.zip` ..... LaTeX thesis source code