Predicting Optimal Insulin Dose for Diabetes Inpatients by using Reinforcement Learning

Dimosthenis Sotiriou

October 19, 2023

University of the Aegean

School of Engineering

Department of Information and Communication Systems Engineering

Predicting Optimal Insulin Dose for Diabetes Inpatients by using Reinforcement Learning

Master's Thesis

To fulfil the requirements for the Master's Degree in

Internet of Things Environments in Next Generation Networks

at the University of the Aegean under the supervision of

Assoc. Prof. Panagiotis Symeonidis

Information and Communication Systems Engineering, University of the Aegean

Dimosthenis Sotiriou

October, 2023

Contents

List of Tables

List of Figures

List of equations

List of Acronyms

ANN	Artificial Neural Networks
AI	Artificial Intelligence
CGMs	Continuous Glucose Monitors
DSS	Decision Support Systems
DQL	Deep Q-Learning
DRL	Deep Reinforcement Learning
EMR	Electronic Medical Record
ICU	Insulin Sensitivity Factor
IQR	Interquartile Range
ML	Machine Learning
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
MDPs	Markov Decision Processes
MIMIC III	Medical Information Mart for Intensive Care III
MSE	Mean Squared Error
MLR	Multiple Linear Regression
PCC	Pearson Correlation Coefficient
RL	Reinforcement Learning
R2	R-squared (Coefficient of Determination)
RF	Random Forest
ReLU	Rectified Linear Unit
RMAE	Root Mean Squared Error
SVM	Support Vector Machine
STD	Standard Deviation
T1DM	Diabetes Mellitus Type-1
T2DM	Diabetes Mellitus Type-2
WHO	World Health Organisation

Acknowledgments

I express my gratitude to my supervisor, Prof. Panagiotis Symeonidis,
I appreciate the advice and mentoring provided throughout this project.

I am deeply grateful to my beloved Rea,
your love has been a constant source of strength.

to my parents, Kerasina and Sotiris,
for their contagious passion for learning and growth.

This accomplishment is as much yours as mine.

για τον Χρήστο, τον αγαπημένο μας αδερφό

Abstract

This thesis addresses the critical challenge of predicting future insulin dosages for both diabetic and non-diabetic patients in Intensive Care Units (ICUs), a setting where precise glycemic control is crucial. The work proposes a unique methodology that sidesteps the complications associated with traditional glucose-insulin interaction models, circumvents the issue of scarce historical patient data, and minimizes the noise introduced by using datasets aggregated from multiple patients. We introduce a simplified version of Bergman's Insulin-Glucose Interaction Model and construct an expansive dataset based on the MIMIC III database. This dataset includes 870 predictive features encompassing demographic data, prior insulin administrations, and average glucose levels. The thesis also introduces a Reinforcement Learning approach, utilizing Deep Q-Learning, to optimize both the instance and the training population selection for individualized predictions.

We found that our Composite Multilinear Regression Models outperformed Single-Patient Regression Models in terms of Mean Absolute Error (MAE) for different demographic groups, including Non-Diabetic and Type II Diabetic patients. Specifically, the MAE values for the Non-Diabetic and Type II Diabetic groups were 2.33 and 3.68, respectively, significantly better than the Single-Patient Models. The work contributes a novel approach to insulin dose prediction, offering a promising pathway for more effective glucose management in ICU settings.

Keywords – Deep Q-Learning, Diabetic Patient, Glucose Levels, Insulin Dosing, Multilinear Regression, Reinforcement Learning

1 Introduction

1.1 Background and Motivation

Glucose levels are crucial for the proper functioning of the body, since glucose is the main source of energy for all cells. Maintaining optimal glucose levels is essential for energy production in the body and brain function [9]. Additionally, poorly controlled glucose levels in hospitalised patients can lead to serious morbidity and mortality.

Glucose levels could be in the normal range, consistently higher (hyperglycemia), or drop below (hypoglycemia). For most healthy individuals, normal blood sugar levels are in the range of 80 to 130 mg/dL when fasting (preprandial) and up to 180 mg/dL 2 hours after eating (postprandial). Additionally, for people with diabetes, the American Diabetes Association (ADA) recommends preprandial blood sugar between 100 - 180 mg/dL, postprandial less than 180 mg/dL and 110 - 200 mg/dL for bedtime glucose level [33]. However, the desired range for glucose levels may be different depending on the individual's age, any additional health problems they have, and other factors.

In relation to patients, they can be categorised as either *Non-Diabetic* or *Diabetic*. Diabetic patients can be further divided into Type 1 and Type 2. *Diabetes Mellitus Type 1* (T1DM) is an autoimmune disease in which the body's immune system attacks and destroys insulin-producing cells in the pancreas. Consequently, individuals with T1DM need insulin injections or an insulin pump to regulate their blood sugar levels. On the other hand, *Diabetes Mellitus Type 2* (T2DM) is a metabolic disorder marked by insulin resistance and a relative insulin deficiency. In cases of T2DM, the body cannot produce sufficient insulin or cannot use insulin effectively, causing higher blood sugar levels. Management of T2DM typically involves lifestyle modifications such as diet and exercise, complemented by oral medications or insulin injections.[31]. In particular, more than 95% of diabetic individuals have Type 2 diabetes [40].

Insulin administration is the main tool to maintain glucose levels within the desired range in diabetic patients. Non-diabetic individuals typically do not need such interventions, as their bodies effectively regulate blood glucose levels. In exceptional circumstances, insulin can be administered due to irregular blood sugar levels due to severe illness, injury, stress, steroid-induced medication, or total parenteral nutrition (TPN) administration. Diabetic patients often require regular monitoring of blood sugar, adherence to a balanced diet, exercise, and insulin therapy to manage their condition. In both cases, current glucose control is highly dependent on expert knowledge, which can result in high variability and often suboptimal blood sugars [31], especially in inpatient glucose control.

1.2 Problem Statement

In Intensive Care Unit (ICU) settings, maintaining tight glycaemic control is crucial. Glucose-insulin interaction models can help clinicians determine the optimal insulin infusion rate for critically ill patients to prevent the adverse outcomes of hyperglycemia and hypoglycemia. Glucose-Insulin Interaction Models are mathematical representations that describe the dynamics between glucose and insulin within the human body that encapsulates the underlying physiological processes [30]. In this study, we tackle the task of forecasting future insulin doses for both diabetic and non-diabetic patients entering the ICU for the first time, employing a simple Glucose-Insulin Interaction Model, which helps us avoid the need for any previous patient history, except basic demographic parameters such as sex, age, weight, and whether they were diabetic or not.

The way in which we formulate the problem makes our work an extension of Liu's 2020 work on predicting inpatient glucose levels and insulin dosing by Machine Learning (ML) on Electronic Health Records (EHR) [31]. In this work, the problem of insulin dose prediction is treated as a typical machine learning problem in which historical data is collected, preprocessed, and various models are trained and evaluated. The main problem with this approach is that data taken directly from the patient's history may produce poor results due to scarsity, whereas data from other patients produce poor results due to differences in glucose-insulin interaction between patients.

Consequently, in this thesis, we address the issue of creating a regression methodology to predict future insulin dosages, avoiding the complexities of glucose-insulin interaction models, the scarcity of examining patient historical data, and the noise in predictions generated by datasets produced by multiple patients.

1.3 Methodology and Research Objectives

To achieve this goal, we propose and evaluate a methodology based on the successful design and implementation of the following building blocks.

1. The definition of a simplified version of Bergman's Insulin-Glucose Interaction Model [11] based on Eriksen's work[15], which can be experimentally structured as a Multilinear Regression Model.

2. Definition of a comprehensive set of 870 predictive features that encompass demographic data, previous insulin administration, and average glucose levels.

3. The definition of a set of target variables that describe future insulin administrations and average glucose levels.

4. The construction of a data set based on MIMIC III [23], a popular Electronic Health Records database, which contains the predefined set of predictive and target variables over a large number of insulin administrations and demographically diverse set of patients.

5. The definition and experimental evaluation of a feature selection methodology based on Pearson's correlation coefficient (PCC) between each feature and a target variable among various experimental settings.

6. Evaluation of various subsets of our experimentation dataset with randomly selected, demographically identical, and similar patients. For this work, demographic information is added to the experimentation dataset and a methodology to quantify similarity between two patients is proposed which is based on the Euclidean distance between the Multilinear Regression Model coefficint of two examining patients.

7. The proposition and evaluation of a Reinforcement Learning technique using Deep Q-Learning to tackle both instance and training population selection for individualized predictions.

1.4 Significance of the Study

The significance of this study lies in its innovative approach to address the challenge of predicting future insulin dosages avoiding the complexities of glucose-insulin interaction models, the scarcity of patient's historical data, and the noise in predictions by multiple patient datasets.
Upon the successful conclusion of this research, we will have:

1. The implementation of a methodology for predicting optimal insulin dosages utilizes multiple Multilinear Regression Models. This methodology is rigorously tested and evaluated under various feature sets and instance sets scenarios.

2. Constructed an extensive dataset that encompasses 870 predictive features, drawn from a wide array of insulin administrations and a demographically diverse patient population.

3. The proposition and assessment of a reinforcement-learning strategy, which employs deep Q-learning, focus on optimizing both instance and training population selection for individualised insulin dosage predictions. This diverges from the prevailing trend in current research that mainly focusses on the direct prediction of insulin dosages.

1.5 Thesis Structure

The structure of this thesis encompasses a literature review to contextualise the surrounding work, an introduction to our methodology, a presentation of our experimental findings, and a concluding analysis.

Chapter 2 begins with an overview of diabetes and insulin administration, transitions to a review of the existing literature on the application of machine learning and reinforcement learning in diabetes management, and concludes with a brief overview of electronic health records databases.

Chapter 3 comprises a section on problem formulation, an overview of the essential concepts and tools used in our research, and culminates with an in-depth description of our innovative methodology. Specifically, the methodology introduces the generation of Composite Multilinear Regression Models using Deep Q-Learning for the purpose of predicting optimal insulin dosages.

Chapter 4 presents a comprehensive set of experiments and evaluations, affirming that our methodology outperforms simple linear regression in predicting optimal insulin dosages. The chapter includes a section on constructing our experimentation dataset, the presentation of a toy example to illustrate a simplified version of our approach, a learning curve analysis across diverse training dataset types and sizes, a sensitivity analysis focussing on the Pearson Coefficient Threshold, and finally, an assessment of the performance of the Composite Multilinear Regression Models generated by our methodology.

Chapter 5 includes the derived conclusions and discusses how the results currently achieved can be improved in future academic endeavours.

2 Background Literature

Insulin administration constitutes an essential tool in the management of diabetes and in the regulation of glucose levels in general. This brief review of the literature aims to provide a framework within which the present study is situated.

The content of this chapter is divided into two parts. In the first section, an overview of diabetes treatment through insulin administration is presented. Glucose-Insulin interaction models such as Bergman's Minimal Model are presented, followed by insulin injection techniques, insulin types, and administration strategies. In the second section, a review of how machine learning techniques can be employed to deal with the examined problem is provided. Examples of work employing linear regression and reinforcement learning in predicting glucose levels and insulin dosing are presented.

2.1 Overview of Diabetes and Insulin Administration

2.1.1 Glucose-Insulin Interaction Models

Insulin is the main hormone that controls glucose metabolism by signaling fat cells and liver cells to absorb glucose. Due to this fact, insulin administration is the primary tool to artificially maintain glucose levels within the desired range[45]. Glucose-Insulin Interaction Models are mathematical representations that describe the dynamics between glucose and insulin within the human body that encapsulates the underlying physiological processes [30]. The complexity of these models can vary, from simple linear relationships to more complex models that omit various factors such as insulin sensitivity, glucose production, and utilisation rates. Models are developed to understand the intricate balance and feedback mechanisms between glucose and insulin and play a vital role in the research, diagnosis, and treatment of diabetes and other metabolic diseases [37].

The Bergman Minimal Model [11] is a mathematical model designed to describe the Glucose-Insulin regulation system in the body. The model consists of three differential equations that describe the dynamics of glucose and insulin concentrations in plasma.

The key equations are as follows:

Bergman's Minimal Model

$$\begin{aligned}\frac{dG}{dt} &= -p_1 \cdot G(t) - X(t) \cdot G(t) + p_1 \cdot G_b + R_a(t) \\ \frac{dX}{dt} &= -p_2 \cdot X(t) + p_3 \cdot I(t-\tau) \cdot (G(t) - G_b) \\ \frac{dI}{dt} &= -n \cdot (I(t) - I_b) + \frac{u(t)}{V_i}\end{aligned} \tag{1}$$

where:

- $G(t)$ is the plasma Glucose concentration at time t
- $I(t)$ is the plasma Insulin concentration at time t
- $X(t)$ is the remote Insulin-dependent glucose uptake at time t
- G_b and I_b are the basal (fasting) glucose and insulin concentrations, respectively
- p_1, p_2, p_3 are rate constants
- $R_a(t)$ is the rate of appearance of glucose.
- n is a rate constant for insulin disappearance
- τ is a time delay
- $u(t)$ is the insulin infusion rate.
- V_i is the distribution volume of Insulin

The minimal model focusses on the dynamics between glucose and insulin, ignoring many other complexities of the metabolic system. Due to this fact, it is a simplified but powerful way to study and understand the glucose-insulin interaction, with valuable applications in the field of diabetes research and treatment.

2.1.2 Insulin Administration Strategies

The goal of insulin therapy is to mimic the physiological patterns of insulin secretion in the body as closely as possible to achieve optimal blood glucose control and to reduce the risk of short- and long-term complications. Insulin administration requires careful monitoring of blood glucose levels and a thorough understanding of diabetes management for each specific patient. Despite this fact, there exist general strategies which can assist in the above goal. The Total Daily Insulin Requirement (TDI) represents the total amount of exogenous insulin a person with diabetes needs in a 24-hour period to maintain blood glucose levels within the target range [13].

TDI Estimation based on Body Weight During the beginning of insulin therapy (for example, during the first hours of patient hospitalisation) when the patient's physiology and insulin resistance are

unknown, the patient's weight plays an important role in the calculation of TDI. Understanding the relationship between weight and insulin needs of the patient is crucial for personalised diabetes care. Generally, TDI can be estimated as a function of body weight, often ranging from 0.4 to 1.0 units of insulin per kilogramme of body weight per day for people without significant insulin resistance [25]. The specific ratio depends on factors such as type of diabetes, stage of the disease, and the presence of other comorbid conditions. In overweight or obese patients or those with insulin resistance, the requirement may be higher.

In general, we can devide the TDI dosage into the following 4 categories [20]:

- 0.3 units/kg/day for lean, elderly patients, or patients with risk of hypoglycemia or insulin-sensitive patients
- 0.4 units/kg/day for patients with normal weight
- 0.5 units/kg/day for overweight patients
- 0.6 units/kg/day for obese patients or patients on high-dose steroids or insulin-resistant

TDI can rise during stressful circumstances, possibly reaching 2 units/kg/dayduring episodes of intense stress[12].

Basal Bolus Insulin Therapy employs a combination of *Long-acting Insulin* (basal) administered once or twice a day to maintain a steady level of insulin throughout the day, while *Rapid-acting Insulin* insulin (bolus) is administered before meals to counteract the increase in blood glucose levels due to food intake. It is commonly used in Type 1 diabetes [46]. The ADVANCE study, published in the New England Journal of Medicine, demonstrated that a basal-bolus insulin regimen was associated with a significant reduction in major cardiovascular events in patients with type 1 and type 2 diabetes compared to conventional therapy. The study also demonstrated that it is possible to achieve tight levels of glycaemic control safely using conventional methods [6]. Bolus insulin constitutes approximately 50 to 60% of the total daily insulin dose, which encompasses both carbohydrate coverage and correction of high blood glucose levels. The remaining 40 to 50% of the daily insulin dosage is designated for overnight insulin replacement, fasting periods, and intervals between meals. The prescription for food coverage through bolus dosing is often expressed in terms of the insulin-carbohydrate ratio (I:C). This ratio represents the number of grams of carbohydrate that can be metabolised or eliminated by one unit of insulin. Typically, one unit of rapid-acting insulin is sufficient to process 12-15 grams of carbohydrate, although this range can fluctuate between 4-30 grams or even more, depending on an individual's responsiveness to insulin. Insulin sensitivity, a critical factor in this context, can vary with increasing time of day, differ between individuals, and be influenced by physical activity and stress. Another essential concept is the Insulin Sensitivity Factor (ISF), also known as the High Blood Glucose Correction Factor. The ISF delineates the extent to which one unit of rapid-acting insulin will

reduce blood glucose levels. Generally, a single unit of insulin is required to decrease blood glucose by 50 mg/dl, but this effect can range from 15 to 100 mg / dl or more, depending on personal insulin sensitivities and other conditions [13]. In general, basal insulin requirements are consistent from day to day and can be adjusted according to factors such as activity levels or illness, contrary to bolus insulin doses, which are adjusted at each meal according to the carbohydrate content of the food and the blood glucose level before meal.

Sliding Scale Insulin (SSI) is often used in hospitals since the 1930s or other inpatient settings to manage blood glucose levels in diabetic patients. Administration follows a predetermined scale that specifies the insulin dose corresponding to different blood glucose levels and includes rapid or short-acting insulin. The sliding scale approach offers flexibility and can be effective in controlling post-prandial hyperglycemia. Despite that, it is a reactive strategy and is not effective in glucose control because it treats hyperglycemia after it had already occurred [38].

Fixed, conventional Insulin therapy involves the administration of a combination of short- or rapid-acting insulin and intermediate-acting insulin) twice daily before meals. This regimen is simpler than basal-bolus or pump therapy and may be suitable for individuals with predictable daily routines and eating habits [35]. However, it offers less flexibility in terms of meal timing and content and may result in less optimal blood glucose control.

In *Insulin Pump Therapy* rapid-acting insulin is continuously infused through a small pump. The amount is adjusted to provide different 'basal' rates to meet the varying insulin needs at different times of day. The user can also deliver bolus doses of insulin at mealtimes or to correct for high blood glucose levels. Insulin pump therapy can offer precise control over insulin delivery. Often referred to as an artificial pancreas, hybrid closed-loop systems offer continuous glucose monitoring (CGM) in conjunction with an insulin pump to automatically adjust insulin delivery based on glucose levels [10]. The "hybrid" part of the term means that the system still requires careful management by the user and regular follow-up with healthcare providers to optimise settings.

2.1.3 Overview of Insulin Types

A variaty of insulin types, with different action profiles, have been produced to mimics the body's natural pattern of insulin release more closely [45].

Rapid Acting Insulin, often referred to as mealtime insulin, is a type of insulin that starts to work 15 minutes after injection. Its peak glucose-lowering effect occurs approximately one to three hours after administration, and its effect can last for three to five hours. Due to the fact that it works quickly, it is usually taken just before or even after meals to counteract the increase in blood glucose after eating.

An example brand name for such type of insulin is *Humalog* [2].

Regular Insulin, often referred to as short-acting insulin, is a type of human insulin with a slower onset and longer duration of action compared to rapid-acting insulin analogues. Regular insulin begins to lower blood glucose levels within 30 minutes after subcutaneous injection, reaches its maximum glucose-lowering effect 2 to 3 hours after injection, and has a total duration of action of approximately 6 to 8 hours. Example brands of regular insulin are *Humulin R* [3] and *Novolin R* [7]. It is often used to control blood sugar levels during the period between meals and throughout the night, acting as a bridge for the longer-acting basal insulins.

Intermediate Acting Insulin, also known as *neutral protamine Hagedorn* (NPH) insulin, plays a vital role in the management of blood glucose levels in individuals with diabetes. This type of insulin typically begins to lower blood glucose levels within 2 to 4 hours after injection, reaches its peak glucose lowering effect between 4 and 12 hours, and its total duration of action extends to approximately 18 to 24 hours. An example brand name for this type of insulin is *Humulin N* [17]. Intermediate-acting insulin contains protamine which slows insulin absorption, allows for less frequent dosing, usually once or twice daily, and provides a steady level of insulin over a longer period.

Long Acting Insulin, also known as basal insulin, serves as the foundation of insulin therapy for most people with diabetes. This type of insulin is designed to be released slowly, providing a steady, continuous level of insulin over a 24-hour period, with minimal to no pronounced peak. This slow and steady release closely mimics basal insulin secretion by a healthy pancreas and serves to control glucose levels between meals and overnight. An example brand name for such type of insulin is *Glargine Lantus* [4].

2.1.4 Overview of Insulin Administration Methods

The methodology for insulin administration varies and is typically dependent on several factors, including the type of diabetes diagnosed, the lifestyle of the individual, and a comprehensive evaluation of the advantages and disadvantages of each method. The primary method of administration is subcutaneous injection, where insulin is administered into the adipose tissue located beneath the skin. This form of administration can be carried out in a sporadic manner using devices such as syringes and insulin pens, or it can be administered continuously with the aid of an insulin pump. These devices allow a regulated amount of insulin to be injected into the body, offering flexibility and control over blood sugar levels [42].

An alternative to traditional needle-based insulin administration is inhaled insulin. This method involves the administration of insulin through the mouth and through the lungs. This form of delivery is

particularly beneficial for people who have needle phobia or require frequent injections, although it is used in conjunction with long-acting insulin injections to achieve complete control of diabetes. Other less commonly used, but emerging options for insulin delivery include insulin patches, implants, and insulin jet injectors. Insulin patches and implants are designed to release insulin through the skin or internally for a specified duration, with the aim of offering a painless and more convenient insulin delivery option. On the other hand, insulin jet injectors use high pressure to deliver a fine spray of insulin through the skin, eliminating the use of a needle.

2.2 Machine Learning in Diabetes Management

Machine Learning (ML) is a subset of Artificial Intelligence (AI) dedicated to the creation of algorithms that allow computers to learn, predict, and make decisions without direct programming [5]. Within diabetes management research, the application of ML has grown and became an essential instrument within the expansive domain of Decision Support Systems (DSS) for patients with Type 1 and Type 2 diabetes. To illustrate the direction of research, we can pinpoint three research tasks and present equivalent examples of past work.

- *Glycemic Level Prediction*, which describes the task of predicting blood sugar levels based on historical data and other patient factors,
- *Insulin Dose Prediction*, the task of predicting the amount of insulin a patient will need at a given time,
- *Diabetes Diagnosis and Risk Assessment*, which assess the existance of diabetes based on patient parameters.

2.2.1 Inpatient Glucose Levels and Insulin Dosing by Liu et al.

A study on predicting inpatient blood sugar levels and insulin dosages using supervised machine learning techniques on electronic health record data was contacted by [31]. The study focused on selecting the most appropriate features from Stanford Research Repository Datalake using Pearson's correlation, applying the Support Vector Regression (SVR) and Tree-Based Random Forest Regression Algorithm (RF), and evaluating the results using the Mean Absolute Error (MAE) and the R2 Score (coefficient of determination) with 95% Confidence Interval (CI).

Specifically, the predicting variables were the current glucose level, the average glucose level in the next 24 hours, and the insulin ordered in the next 24 hours. The data set used was a STARR subset of 42.700 patients with glucose levels measured either $\geq$ 200 mg/dL or $\leq$ 70 mg/dL. The data was also filtered to patients with prescribed insulin, recorded weight, at least five glucose measures within 72 hours, and without hemodialysis treatment. The resulting data set contains 3,461 patients and

175.934 unique data points. Feature selection was performed from a list of 20 initial features using Pearson's correlation. The experimental results showed that by feature selection the run time of the algorithms was reduced, and the prediction quality was also slightly reduced. The final list of selected features includes the average glucose level in the past 24 hours, the glucose level at a similar time as the previous day, the variance of the glucose level in the last 24 hours and the current glucose level. The results of the study show that individual blood glucose level and insulin dosing are in an erratic end, and accurate prediction is impractical. The average blood glucose level over 24 hours can be more reliably predicted.

2.2.2 Predicting Initial Inpatient Total Daily Dose by Nguyen et al.

A different study [34] investigates whether machine learning can provide more precise predictions for the initial total daily dose (TDD) of insulin in the hospital based on electronic health records compared to current dosing recommendations based on weight-based guidelines. The work consists of two distinct experiments: 1. Predicting, as a binary outcome, whether a patient will need more than 6 units of TDD, distinguishing between *low* and *higher* requirements, and 2. for patients who require more than 6 units of TDD, predicting the specific TDD value necessary to achieve proper glucose control. The research used electronic health records from a tertiary academic centre spanning from 2008 to 2020. These records belonged to 16.848 patients who received subcutaneous insulin and achieved a target blood glucose control range of 100-180 mg/dL in at least three blood glucose measurements in a single day.

The chosen threshold of 6 TDD units was based on the fact that approximately 75% of the patients needed 6 units or fewer. Machine learning algorithms used included regularised regression, random forest, and gradient-boosted tree, and incorporated 87 features such as weight, height, age, sex, race, insurance status, creatinine levels, diet, microbiology lab order counts and glucocorticoid usage. The evaluation methods involved the area under the receiver operating characteristic curve (AUROC) and the area under the precision recall curve (AUPRC).

In the comparison of low versus higher TDD, a weight-based only classifier did not perform much better than random chance (AUROC 0.57, AUPRC 0.29). However, when using the full set of variables readily accessible in electronic medical records, the machine learning classification approach showed a significant improvement (AUROC 0.85, AUPRC 0.65), highlighting its ability to differentiate between low and high insulin users. For the more challenging task of determining the precise TDD value of points for higher users, the generalised linear regression model outperformed both the random forest and gradient-boosted tree models.

2.2.3 Diabetes Prediction using Machine Learning by Suresh et al.

The work of [28] used the PIMA Indian database from the UCI repository to predict diabetes. The dataset, consisting of 768 entries with eight input features and one target variable, was analysed using R-Studio. The research employed multiple ML algorithms, including Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Networks (ANN). Feature selection identified age, BMI, and glucose level as vital to the prediction. The SVM model was the most accurate with an 80.3% success rate, highlighting age, BMI, and glucose level as the primary prediction indicators. The authors suggest that validations with larger datasets and recent ML techniques could improve prediction accuracy.

2.3 Reinforcement Learning in Diabetes Management

2.3.1 Definition of Reinforcement Learning

Reinforcement Learning (RL) comprises algorithms designed to tackle Markov Decision Processes (MDPs). MDPs are structures used to depict decision-making in scenarios with unpredictable results, influenced by both randomness and a decision maker's actions. Essentially, an MDP guides an agent in optimal decision making by weighing immediate and future rewards [24]. In RL, the surrounding environment is typically represented as a Markov chain. Here, the state of the subsequent environment is based only on the present state, disregarding previous states, reducing the problem solving process but potentially oversimplifying real world learning situations [8]. RL focusses on training agents to make optimal decisions by interacting with their environment. These agents learn to maximise the cumulative reward, which represents the objective and purpose of building the agent, by mapping states to actions through a trial-and-error process. This process involves balancing exploration, where the agent tries new actions to gather information about the environment, and exploitation, where the agent leverages its current knowledge to select the best known action [29].

The fundamental components of a reinforcement learning problem are as follows:

- *Agent*. The entity that is being trained to make decisions in the environment.
- *Environment*. The context where the agent operates, including any constraints or laws that govern state transitions and rewards.
- *State*. A representation of a distinct situation within the environment.
- *Action*. A decision made by the agent that affects its state or the environment.
- *Reward*. Feedback provided by the environment indicating the desirability of the agent's actions.

- *Policy*. A strategy used by the agent to map states to actions that guide its decision-making process.

The primary goal of reinforcement learning is to find an optimal policy that maximises the expected cumulative reward over time.

2.3.2 Introduction to Q-Learning

Q-Learning is a model-free reinforcement learning algorithm used to determine the optimal action selection policy for a given finite Markov decision process. In other words, it helps agents learn how to choose optimal actions that yield the most reward over time, even when they don't know anything about the environment's transition probabilities. The Q in Q-learning stands for *quality*, which represents the value of a particular action in a given state [43]. The main idea behind Q-learning is to estimate the value for every pair of state actions (s,a), and this value is represented as $Q(s,a)$. An agent then uses these Q values to make decisions. The value of $Q(s,a)$ is the expected return from taking action a in state s and then following the optimal policy [47]. The core of the Q-learning algorithm involves updating the Q-values using the Bellman equation as follows:

Bellman Equation

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\left(r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)\right) \tag{2}$$

Where:

- s_t and a_t are the current state and action, respectively.
- α is the learning rate, which determines the extent to which new Q values are taken over the old ones.
- r_{t+1} is the reward received after taking action a_t in the state s_t.
- γ is the discount factor that models the agent's consideration for future rewards. It ranges from 0 to 1.
- $\max_a Q(s_{t+1},a)$ represents the maximum Q-value for the next state s_{t+1}, for all possible actions a.

Over time, as the agent interacts more with its environment and receives feedback in the form of rewards, the Q values approach the true values, and the agent becomes better at selecting actions that maximise its rewards.

2.3.3 Deep Reinforcement Learning and Deep Q-Learning

Deep Reinforcement Learning (DRL) is a specific kind of Reinforcement Learning that uses Deep Neural Networks (DNN) as function approximators with Reinforcement Learning methodology. Due to the fact that DNN can handle large-scale, high-dimensional state and action spaces, DRL is suitable for solving more complex problems than traditional reinforcement learning methods [8].

Deep Q-Learning (DQL) is a model-free, online, off-policy reinforcement learning method. The primary goal in Q-Learning is to learn a policy that acts optimally, maximising the expected cumulative discounted reward. In Deep Q-Learning, the Q-values are approximated using a neural network. Therefore, the update equation becomes a loss minimization problem:

Temporal Difference Error for Q-Learning

$$L(\theta) = E_{(s,a,r,s')\sim U(D)}\left[\left(r + \gamma \max_{a'} Q(s', a'; \theta^{-}) - Q(s, a; \theta)\right)^2\right] \tag{3}$$

Where:

- θ are the parameters of the Q-network.
- D is a replay buffer containing experience tuples (s, a, r, s').
- (U (D)) denotes the uniform sampling of the replay buffer.
- θ^{-} are the parameters of a target network, which are periodically updated with the parameters of the Q network. This target network helps stabilise the learning process.

The neural network is trained to minimise this loss, allowing the model to learn the optimal Q values over time [18].

2.3.4 Brief Review of Reinforcement Learning in Diabetes Management

According to Tejedor [44], the field of RL is expanding rapidly, especially in its application to the regulation of diabetes. The review of literature reveals a surge in interest: Before 2012, there were only two related publications, but between 2012 and 2019, this number increased to 27.

In these studies, Actor-critic (AC) was the most popular RL algorithm (36.67%) followed by Q-learning (10%). The most popular data source is UVA/PADOVA Type 1 Diabetes Simulator, a virtual patient population with various demographic and physiological characteristics, approved by the U.S. Food and Drug Administration (FDA)[32]. Real data was used in only 20% of the studies. Bergman's

minimal model was used as a data source in 12.90% of the cases[44]. In case of the Exploitation-Exploration dilemma 24.25% of the times Gaussian noise was added to the agent's action values to choose the non-optimal solution, while -greedy policies were used in 12.12%.

Regarding the state space, 73,33% of the times continuous variables were used and the blood glucose level the most popular, followed by the previous insulin dose, the carbohydrate intake (CHO) and the weight. Regarding the action space, continuous variables were again the most frequent and insulin dose the most frequent. This was followed by a small minority that aimed at blood glucose and food intake. The reward function, which is crucial for the design of an RL system, showed great variability. In the vast majority of cases, a reward was given when the output measurement was within the normal predefined range.

2.4 Electronic Health Records

Electronic Health Records databases contain the history of hospital encounters, records of diagnoses and interventions, lab tests, medical images, and clinical narratives. All these datasets can be used to build predictive models that can help clinicians with diagnostics and various treatment decision support. Example of such datasets are those of PCORnet, the National Patient-Centered Clinical Research Network [39]. PCORnet is an innovative initiative designed to improve the nation's capacity to conduct clinical research by creating a large and highly representative network for conducting clinical outcome research. PCORnet datasets are typically not available to the public due to the sensitive nature of patient data and the strict regulations surrounding health information [16].

In diabetes management, the necessary data encompass glucose levels, historical insulin doses, food consumption information, physical activity, and various demographic details. These data can be obtained from continuous glucose monitors (CGM), insulin pumps, food tracking applications, and activity tracking devices.

2.4.1 MIMIC-III Dataset

Medical Information Mart for Intensive Care III (MIMIC III) v1.4. [23] is a freely available database that contains health related data of 46.500 patients of Beth Israel Deaconess Medical Centre between those who stayed between 2001 and 2012. The database includes vital sign measurements, laboratory tests, procedures, medications, and caregiver notes. A successor to MIMIC II v2.6 includes an additional 20000 additional ICU admissions, physician progress notes, medication administration data, and more complete demographic information. The dataset consists of tables stored in csv files, from which 21 of them track patient stays in hospital and ICU, and an additional 5 tables are structured as dictionaries providing definitions for identifiers with in the dataset.

3 Research Methodology

The objective of our study is to explore the potential of predicting insulin dosages for diabetic and non-diabetic ICU patients, using existing electronic health records and machine learning techniques. In this chapter, we outline the methodology proposed to accomplish this objective, also presenting a comprehensive description of our experimental goals, the rationale behind our design decisions and assumptions, the algorithm used, and the evaluation metrics selected for the task.

3.1 Problem Formulation

Estimating total daily insulin (TDI) of a patient who has just entered the intensive care unit and then the individual dose is a complex process that requires careful evaluation of general patient health, stress levels, severity of diabetes, weight, age and other concurrent medical conditions[12]. After the initial estimation of TDI, continuous monitoring of blood glucose levels is required and reevaluation of insulin dose is vital, since insulin requirements can change rapidly in critically ill patients. To give an idea of the complexity, using our experimentation dataset which will be presented in detail in the next chapter, we generated Figure 1 that shows the number of days in the ICU per TDI / Weight Factor.

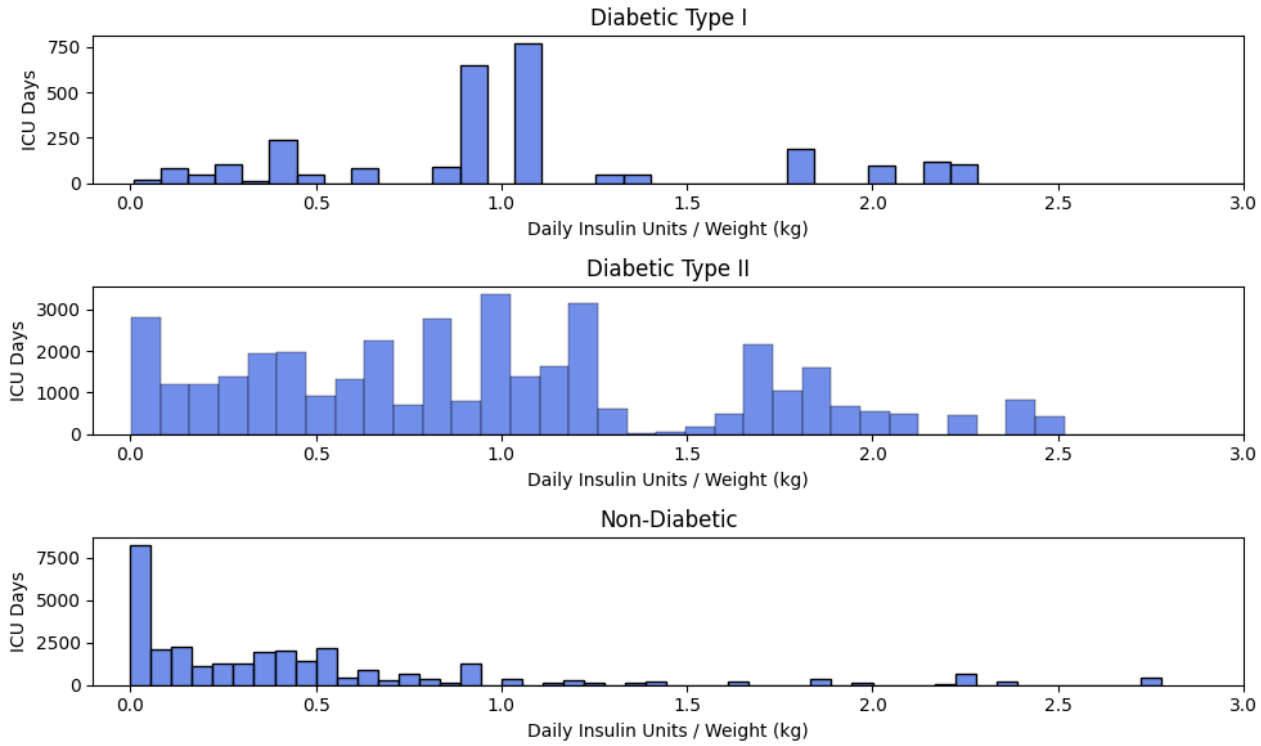

Figure 1: Number of ICU Days per TDI/Weight Factor

Significant differences in the TDI / weight ratio are observed in the overall data set. This suggests that general insulin guidelines might not be suitable for ICU settings and that individualised insulin

dosing may be necessary.

Since a patient's response to insulin is based on the current physiology of the patient, and we target the normal glucose levels presented in Section , we believe that future insulin doses can be predicted from past ones if we can determine the current (relevant to the task) patient's physiology. As a result, the challenge of predicting future insulin doses essentially revolves around collecting enough data on insulin responses in relation to the physiology of the patient. However, when a patient is admitted to a hospital or ICU for the first time, there may not be sufficient prior data, especially if insulin was administered infrequently. Consequently, the problem can be further transformed into determining how to supplement these missing data using insulin response information from other patients with physiology similar to the patient being examined.

3.2 Overview of Fundamental Concepts and Tools

Given the problem as it was formulated in the previous section, before the introduction of our reinforcement learning methodology, it's crucial to outline the key concepts, algorithmic tools that serve as its foundation. Specifically:

- The adoption of a simplified Insulin-Glucose Interaction Model,
- The definition of an Insulin Administration Strategy,
- The definition of a set of dependent variables which describe a) the patient's future glucose levels and b) future insulin doses administered to the patient,
- The definition of a set of independent variables that describe the current state of a patient with respect to insulin administered and current and past glucose levels,
- The definition of a feature selection methodology centred on Pearson's correlation coefficient (PCC) between individual features and a target variable,
- The definition of a feature preprocessing methodology,
- The definition of a training instance selection methodology based on randomly selected, demographically identical and similar patients,
- The definition of a set of evaluation metrics, capable of quantifying the performance of our experintation work.

In the subsequent phase, following the introduction of the above key foundational concepts, we will introduce our approach using Deep Q-Learning to optimise the training dataset population size and feature selection for more personalised predictions.

3.2.1 Insulin-Glucose Interaction Model as Multilinear Regression

The methodology of our experimental work is based on a simplified version of Bergman's minimal model, which was presented in section 2.1.1. The simplified model is based on the work of Eriksen [15] on model-based control for closed-loop insulin delivery systems.

Eriksen presents a linear approximation of Bergman's minimal model which slides over the complex calculations of Bergman's model parameters. It is worth mentioning that this model is too simple to be useful in practise, but it is a good starting point for further experimentations. The models focus on 5-minute period blood glucose change calculations taking into account the amount of insulin absorbed into the blood, digested carbohydrates, and the amount of endogenous glucose production. Mathematically, this is represented by the following equation.

Simplified Bergman's Minimal Model

$$\delta_t = c_e + c_C * C_t - c_I * I_t + c_\delta * \delta_{t-1} \tag{4}$$

where:

- C_t the amount of carbohydates absorbed in period t
- I_t the amount of insulin absorbed in period t
- c_e the amount of endogenous production (independent of t)
- δ_{t-1} change in blood glucose at time $t-1$

The coefficients (c_e, c_C, c_I, and c_δ) can be learnt by the model to weigh the relative importance of the various inputs.

To further simplify the model, during the generation of our experimental dataset, we adjust the blood glucose change calculations to a 15-minute period, reducing the size of the data set and the processing times. In our methodology, we stick to the fundamental principle of linearity with respect to actions and future states, and we substitute the variables in the equation with our ensemble of predictive features.

Since the Simplified Bergman's Minimal Model assumes a linear correlation between dependent and independent variables, we will employ multiple linear regression in our experiments. Multilinear regression is a statistical technique used to model the relationship between a dependent variable and two or more independent variables [21]. Unlike simple linear regression, where only one independent variable is used, multilinear regression incorporates multiple predictors to provide a more comprehensive understanding of the underlying relationship. The mathematical formulation of multilinear regression can be expressed as:

Multilinear Regression Model Equation

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon \tag{5}$$

where Y is the dependent variable, β_0 is the intercept, β_1, β_2, ..., β_p are the coefficients of the independent variables X_1, X_2, ..., X_p, and ε represents the random error term. The coefficients are estimated using methods such as the least squares technique, in order to minimise the sum of the squared differences between the observed values and the values predicted by the model.

3.2.2 Insulin Adminsitration Strategy

In our literature review we mensioned various insulin adminsitration strategy (see Section 2.1.2), including *Sliding Scale Insulin* whch is a simple reactive approach to administer insulin and *Basal Bolus Insulin Therapy* which is a more complex proactive approach which is based on basal administration to maintain a steady level of insulin throughout the day and bolus administrations to counteract the increase in blood glucose levels due to food intake.

Since MIMIC, the primary source of our data, does not include a significant amount of food intake information, our insulin administration strategy adheres to the following rule: Multilinear regression models are trained to predict insulin dosages that maintain glucose levels within the target range of 80 to 180 mg/dL. For instance, if we are targeting the Next-1-Hour Short-Acting Insulin Dosage, we select training examples that predict Next-2-Hours Average Glucose Levels within the range of 80 to 180 mg/dL. This range, is chosen for simplicity to encompass both non-diabetic and diabetic patients.

Given that ICU patients typically have regular feeding schedules, in future work, we may explore filtering training instances based on the time of day.

3.2.3 Definition of Independent and Dependent Variables

We generate a set of prediction attributes and target variables that we intuitively believe capture the current and future state of each patient with respect to their future average glucose level and insulin doses and potentially distinguish them from the broader population.

In case of Independent Variables the derived features are based solely on the patient's average past glucose levels and the cumulative amount of rapid-acting (humalog insulin), short-acting (regular insulin), and long-acting (glargine insulin) administered at a particular pasted time range.

We partitioned the patient's ICU stay into consistent time intervals. The first interval begins after the patient's admission to the ICU, with subsequent intervals followed in such a manner that each glucose

measurement falls exclusively within one interval, and every insulin administration corresponds to at least one interval.

In Eriksen's research on closed-loop insulin delivery systems [15], 5-minute intervals were adopted. In contrast, our empirical study uses 15-minute intervals, taking into account that rapid-, short-, and long-acting insulins have longer durations of action. The duration of these time intervals is defined as *step s* and durations features and target variables are defined as multiples of this step, represented as *period p*.

For a given time interval dt of a patient in ICU we generate the following sets of predictor variables:

- The average Glucose Level during past period p
- Standard Deviation of Glucose Level during past period p
- The total administered amount of all types of Insulin during past period p
- The total administered amount of Rapid-Acting Insulin during past period p
- The total administered amount of Short-Acting Insulin during past period p
- The total administered amount of Long-Acting Insulin during past period p

where p spans the range of 15 minutes to 36 hours, incrementing in steps of 15 minutes or briefly $p \in \{15, 30, 45, \ldots, 129.600\}$ seconds.

Similarly, we define a set of dependent variables that describe a) the patient's future glucose levels and b) future insulin doses administered to the patient. For a given time dt of a patient in ICU we generate the following sets of target variables:

- The average Glucose Level during the following period p
- The total administered amount of all types of Insulin in the following period p
- The total administered amount of Rapid-Acting Insulin in the following period p
- The total administered amount of Short-Acting Insulin in the following period p
- The total administered amount of Long-Acting Insulin in the following period p

p spans the range of 15 minutes to 2 hours, incrementing in steps of 15 minutes or briefly $p \in \{15, 30, 45, \ldots, 7.200\}$ seconds.

3.2.4 Feature Scoring Methodology

In this Section, we describe the selected method for feature selection, which will help us to identify and retain only the most significant variables, and consequently improve our prediction models accuracy. Since we assume linear interaction between administered isnulin and patients' glucose levels, we employ Pearson's correlation coefficient (PCC), in order to quantify the linear association between each attribute and every target. The methodology we are using is a common filtering technique in which features are scored and filtered before the model is built.

The Pearson correlation coefficient, often denoted as r, is a measure that quantifies the linear relationship between two variables. The coefficient's value ranges from -1 to 1, where a value of 1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and a value of 0 suggests no linear relationship between the variables [41]. The formula for the Pearson Correlation Coefficient for a pair of variables X and Y with n data points is given by:

Pearson Correlation Coefficient (PCC)

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \tag{6}$$

where x and y are individual data points of the variables X and Y

Since we were interested in the magnitude of the correlation, we always calculate the absolute value of r. Furthermore, we define a minimum threshold θ_{min} for each independent - dependent variable pair with which we can filter out independent variables that have a weak correlation with the specific target variable, which can negatively affect performance. Through a sensitivity study, adjusting the value of θ_{min}, we aim to determine the optimal threshold for an examining patient at a speicif examination time.

3.2.5 Data Preprocessing

In this section, we present the preprocessing steps undertaken to refine and transform our set of features. The methods applied include handling of missing values and scaling. Insulin administered amounts and glucose measurements have different units and ranges, making scaling essential to bring all features to a common scale, making them directly comparable.

Specifically, we execute the following steps for each feature:

- We calculate the mean value using a random sample of 500 patients
- For columns of Insulin administered amounts, retain values > 0

- Replace missing values with the mean value of the precalculated 500 patients
- Scaling using -score normalisation

Z-score normalisation, commonly known as standardisation, is a statistical method used to standardise and transform a data set such that its mean is 0 and its standard deviation is 1. [19]. The formula for Z-Score Normalisation for a feature x is given by:

Z-score Normalisation

$$x_{scaled} = \frac{x_{initial} - \bar{x}}{\sigma_x} \tag{7}$$

where x_{initial} is the original value, $\bar{x}$ is the mean of the feature, and σ_x is its standard deviation.

3.2.6 Three Techniques for Instance Selection

In our experimental work, instance selection is performed at the level of individual patients. This means that we choose to select a patient based on various criteria, (which will be presented shortly) and then the entire data from the selected patient are included in the training set.

In our experimental work, we employ three methods of instance selection.

- Randomly Selected Patients. In this approach, patients are chosen randomly from the general population of our experimental dataset, without any specific criteria.
- Selection based on the following demographic groups.
  - Based on *Gender*: *Male* and *Female*,
  - Based on *Weight Group*: *Low Weight* (less than 75 kg), *Middle Weight* (from 76 to 100 kg), *High Weight* (from 100 to 120 kg) and *Extreme-High Weight* (more than 120 kg),
  - Based on *Diagnosis*: *Diabetic Type I*, *Diabetic Type II* and *Non Diabetic*,
  - Based on *Age Group*: *Young* (less than 45 years), *Middle-Age* (from 46 to 65 years) and *Old* patients (more than 66 years).

  The clusters for *Weight Group* and *Age Group* were designed to ensure a more even distribution within the experimental dataset, while also maintaining descriptive relevance.
- Based on patient similarity, the similarity is measured using the Euclidean distance between the coefficients of the multilinear model for the patient under examination and the coefficients of the multilinear model for each patient of the rest of the population.

The similarity between two patients is given by the equation:

Similarity between Patients p and q

$$S(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{8}$$

where

- $S(p,q)$ is the Similarity between patient p and patient q.
- p and q are points in n-dimensional space defined by the set of features in Section 3.2.3. $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$.
- p_i and q_i represent the i-th feature of patients p and q respectively.
- n is the number of features as these where defined in 3.2.3.

The total similarity distance is calculated for the entire set of features as defined in Section 3.2.3. For features that are excluded due to feature selection, the value of zero is assigned.

3.2.7 Accuracy Evaluation Metrics

All our evaluations and analyses rely upon the set of three fundamental metrics, which are introduced in detail within this section.

Mean Absolute Error (MAE) measures is a commonly used metric to evaluate the accuracy of regression models employing the average magnitude of errors in a set of predictions. Lower values indicate better predictive accuracy and closer alignment between predictions and actual values.

Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

where:

- n the number of observations in the dataset,
- y_i the true value of the y$^{\text{th}}$ observation,
- $\hat{y}_i$ the predicted value for the y$^{\text{th}}$ observation,

Mean Squared Error (MSE) is a metric used in regression analysis to evaluate model performance. Mathematically, for the observations (n) the MSE is calculated as the average of the squares of the differences between the actual observed results and the predictions made by the model. Formally, for actual values y_i and predicted values $\hat{y}_i$, the MSE is defined as:

Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

To make the interpretations of MAE more interpretable, we can use the *Root Mean Squared Error (RMAE)* which is given by the equation $\text{RMSE} = \sqrt{MSE}$.

The value of (R2), known as the coefficient of determination, serves as a metric that quantifies the proportion of variance in the dependent variable that can be attributed to the independent variables [14]. Its value ranges between 0 and 1, where a higher R^2 means that the model accounts for a greater proportion of the variance in the dependent variable. Mathematically, it is defined as follows:

Coefficient of Determination

$$R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}} \tag{11}$$

where SS_{res} denotes the residual sum of squares and SS_{tot} represents the total sum of squares. An R^2 value closer to 1 indicates that the model provides a good fit to the observed data, capturing most of the variance, while a value closer to 0 suggests the opposite.

3.3 Deep Q-Learning on Composite Multilinear Regression Models Generation

The Single-Patient Regression Model of an examining patient P at a given time t, refers to a Multilinear Regression Model trained, and evaluated using the subset of the individual patient's data points up until time t and predicts one of the target variables of Section 3.2.3.

The Composite Multilinear Regression Model (Multi-Patients Regression Model) for an examining patient P at a specific time t refers to a collection of Multilinear Regression Models. Each of these models corresponds to a Single-Patient Regression Model trained using the complete dataset of a patient similar to the one examined. The criteria for similarity, detailed in Section 3.2.6, are based on the Euclidean distance between the coefficients of the Single-Patient Regression Model for patient P at time t and those of other patients in the broader population. The prediction of the Composite Multilinear Regression Model is derived as the average prediction from its individual model and employs as input the datapoint of the examining patient P at a specific time t. Performance assessment is based on the data points from the examining patient P. A Single-Patient Regression Model qualifies for inclusion in a Composite Multilinear Regression Model when its performance aligns within the specified range: $MAE \geq 0, \quad MAE < 5, \quad R2 \geq 0.2, \quad$ and $\quad R2 < 1$. The evaluation is performed in

a 80% training/20% evaluation using the patient data set.

3.3.1 Prediction Process Using Composite Multilinear Regression Models

According to the definition of Composite Multilinear Regression Models, the process of insulin dose prediction for a patient P at a specific time t involves multiple steps. Initially, a Single-Patient Regression Model is generated for the patient under examination, utilizing data available up to the time point t. Subsequently, our Deep Q-Network is employed to identify an optimal population of similar patients as well as the ideal value for the Pearson Correlation Coefficient (PCC) threshold. Finally, individual regression models are trained for each patient in the set, using their complete datasets. In the final stage, the current state of the patient under examination at time t is fed into each individual model separately, yielding distinct predictions. The Composite Multilinear Regression Model then outputs the average of these individual predictions. A detailed description of this algorithmic process is provided in the Figure 2.

Algorithm 1 Insulin Dose Prediction Process by emplying Composite Multilinear Regression Models in an examining patient P at a given time t

procedure PREDICTINSULINDOSE(t_{now}, patient, patients_list, target)
    // Calculate Single-Patient Regression Model
    $data_{patient} \leftarrow$ getPatientData(patient, Time_limits $= (0, t_{now})$)
    $data_{patient} \leftarrow$ applyPCCthreshold($data_{patient}, \theta = 0.1$)
    $data_{train}$, dataset_eval $\leftarrow$ splitData($80\%, 20\%, data_{patient}$)
    $model_{single}$, $MAE_{single} \leftarrow$ trainEvalModel($data_{train}$, dataset_eval, target)
    $coeff_{single} \leftarrow$ getCoefficients($model_{single}$)
    if $MAE_{single} > 5$ **then**
        return No Prediction
    else
        // Find optimal (population, θ) using DQN
        $population_{optimal}, \theta_{optimal} \leftarrow$ askDQN($coeff_{single}$, population $= 5, \theta = 0.5$)
        // Calculate similarities
        similarities_list $\leftarrow$ EMPTY ARRAY
        for each p in patients_list **do**
            similarity $\leftarrow$ calculateSimilarity(patient, p)
            APPEND *similarities_list*, (*similarity*, p)
        end for
        SORT *similarities_list* by similarity in DESCENDING order
        patients_selected $\leftarrow$ SELECT FIRST$population_{optimal}$ items from*similarities_list*
        // Create Composite Multilinear Regression Model
        models_list $\leftarrow$ EMPTY ARRAY
        for each *patient_sim* in *patients_selected* **do**
            $data_{patient\_sim} \leftarrow$ getPatientData(*patient_sim*, Time_limits $= (0, t_{max})$)
            $data_{patient\_sim} \leftarrow$ applyPCCthreshold($data_{patient\_sim}, \theta = \theta_{optimal}$)
            $data_{train}$, dataset_eval $\leftarrow$ splitData($100\%, 0\%, data_{patient\_sim}$)
            model, MAE $\leftarrow$ trainEvalModel($data_{train}$, dataset_eval, target)
            APPEND *models_list*, *model*
        end for
        // Get examining patient current moment
        $data_{patient\_now} \leftarrow$ getPatientData(patient, Time_limits $= (t_{now}, t_{now})$)
        $data_{patient\_now} \leftarrow$ applyPCCthreshold($data_{patient\_now}, \theta = 0.1$)
        // Calculate Composite Multilinear Regression Model Prediction
        predictions_list $\leftarrow$ EMPTY ARRAY
        for each *model* in *models_list* **do**
            prediction $\leftarrow$ predict(*model*, $data_{patient\_now}$)
            APPEND *predictions_list*, *prediction*
        end for
        $n \leftarrow$ LENGTH(*predictions_list*)
        result $\leftarrow \frac{\text{SUM}(predictions\_list)}{n}$
        return *result*
    end if
end procedure

Figure 2: Insulin Dose Prediction Process by emplying Composite Multilinear Regression Models in an examining patient P at a given time t

3.3.2 Architectural Overview of Reinforcement Learning Framework

The Reinforcement Learning System is a vital component in the process described in the previous section and it was employed as $Population_{optimal}, \theta min_{optimal} = askDQN(coeff_{single}, Population = 5, \theta min = 0.5)$ (see 2, line (12)). As we can see, the component was initialised with random values for $(Population, \theta_{min}$ and returns the optimal values $(Population_{optimal}, \theta min_{optimal})$. To equip the agent with the ability to determine optimal values for both the population and PCC threshold, a reinforcement learning framework is employed. This process is presented in Figure 3.

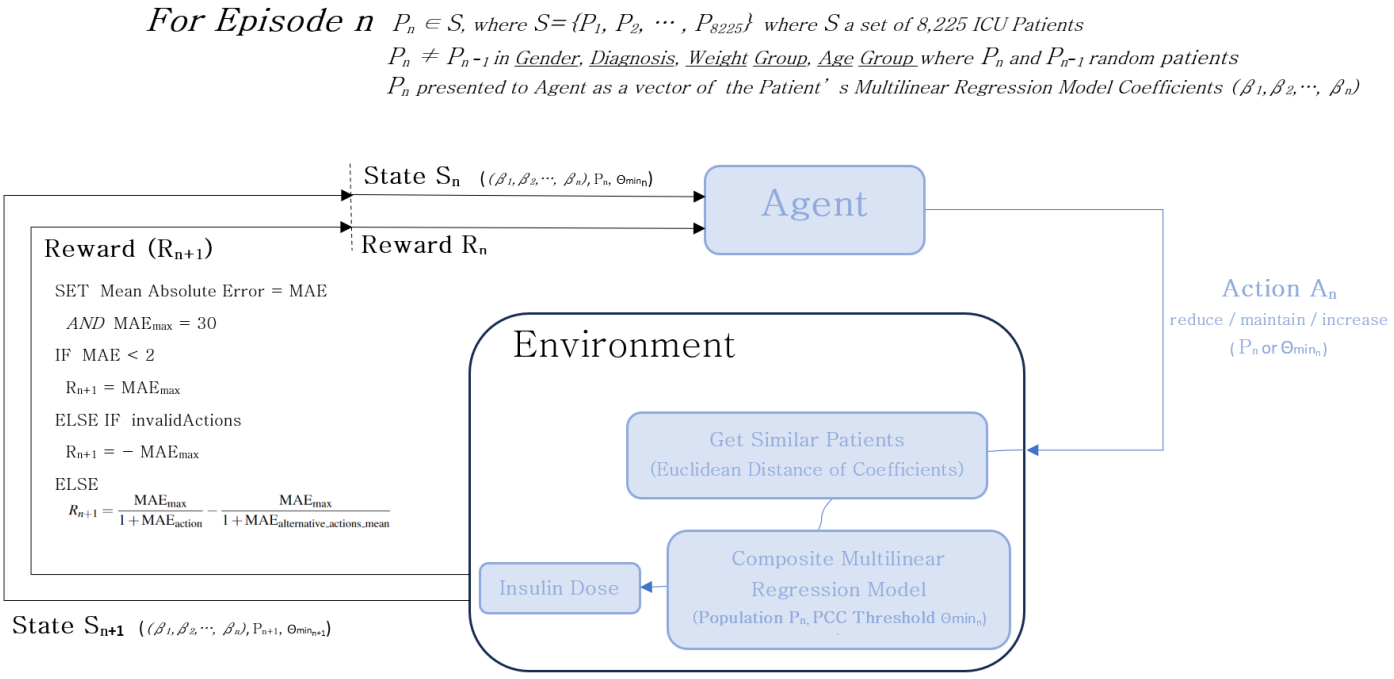

Figure 3: Architectural Overview of the Reinforcement Learning Framework, illustrating the interaction between the Agent, Environment, State, Action, and Reward components

In the following sections, we will outline each of the incorporated components for a comprehensive understanding.

3.3.3 Definition of Policy Function

For the Exploration vs. Exploitation strategy we employed an ε-greedy policy, in which the agent chooses an action uniformly at random With probability ε (Exploration) and selects the action with the highest expected reward (Exploitation) With probability $1 - \varepsilon$ according to the equation 12.

ε-greedy Policy Equation

$$\varepsilon = \varepsilon_{\min} + (\varepsilon_{\max} - \varepsilon_{\min}) \exp(-\text{decay\_rate} \times n) \tag{12}$$

Where:

- ε_{min}. The minimum value of ε, with value 0.01
- ε_{max}. The maximum value of ε, with value 1
- decay_rate. The decay rate of ε, with value 0.0002
- n. The episode number

The Motivation for the ε-greedy policy equation (12) lies in the dynamic nature of exploration and exploitation trade-offs in reinforcement learning. The equation presents a mechanism to control the exploration rate (ε) during the learning process. By gradually decreasing ε over episodes, the agent transitions from a highly explorative strategy to a more exploitative one. This adaptability is essential to find the optimal policy while avoiding getting stuck in suboptimal ones. The equation uses parameters such as ε_{min}, ε_{max}, and decay_rate to fine-tune the exploration-exploitation balance. As the episode number (n) increases, ε diminishes, favouring exploitation over exploration, which can lead to better policy convergence and performance. The chosen decay rate produces an equal probability of exploration and exploitation, approximately at episode 3500.

3.3.4 Reward Function and Evaluation Metrics

The motivation behind the proposed reward function is to provide a quantitative measure that guides the reinforcement learning agent in making decisions that lead to improved performance in optimising two key parameters: the population size within the Composite Multilinear Regression Model and the Pearson Correlation Coefficient (PCC) threshold.

The Reward Function is designed to assess the agent's performance by comparing the Gain of the current action to the average Gain of all alternative actions. The Gain is calculated based on the Mean Absolute Error (MAE) metric obtained from predictions made by the composite multilinear regression model. A critical component of the Reward Function is the introduction of a maximum threshold value, denoted as $MAE_{max} = 30$. The inclusion of MAE_{max} plays a pivotal role in ensuring the stability of the learning process. It serves to mitigate the impact of episodes characterized by exceptionally large prediction errors, preventing extreme outliers from unduly influencing the agent's behavior.

Since higher MAE values indicate poorer performance, we define the Gain as $\frac{MAE_{max}}{1+MAE_{action}}$ for the current action and $\frac{MAE_{max}}{1+MAE_{alternative\_actions\_mean}}$ for the average of alternative actions. Inclusion of the quantity 1 in the denominator ensures that no divisions by zero occur, maintaining the mathematical integrity of the calculation. The difference between the gains is presented in Equation 13

Difference between Gains Equation

$$dG = \frac{\text{MAE}_{\text{max}}}{1 + \text{MAE}_{\text{action}}} - \frac{\text{MAE}_{\text{max}}}{1 + \text{MAE}_{\text{alternative\_actions\_mean}}} \tag{13}$$

The reward function R_n at episode n, contains three parts.

- If the agent selects to maintain both the Threshold and Population constant over two consecutive steps and the current MAE evaluation of the Composite Multilinear Regression Model is less than 2, the agent receives a reward of MAE_{max}.
- If the agent suggests an action that pushes the Population or Threshold beyond the pre-defined limits (see Section 3.3.6), the agent incurs a penalty of -MAE_{max}.
- In all other scenarios, the reward is determined by the equation 13

A graphical representation of this is given in Figure 4.

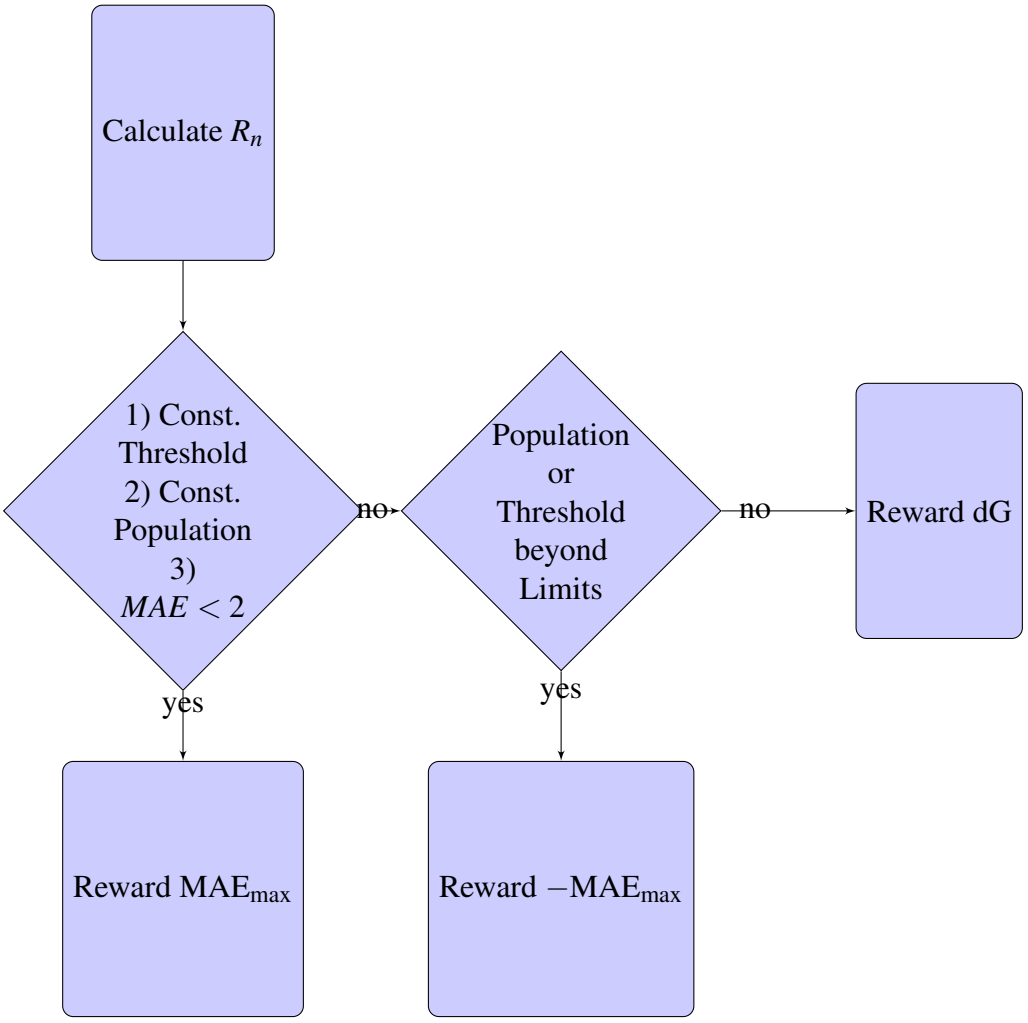

Figure 4: Flowchart for calculating Reward Function R_n at Episode n

3.3.5 Design Characteristics of Reinforcement Learning Environment

The State Space of our designed environment, serving also as the input to the Q-Network, is segmented into three distinct components (Figure 5):

- A vector of coefficients taken from the Single-Patient Regression Model for the patient under examiniation, derived using a subset of features and instances from that of available patient's data. In our experiments, we selected the first 33%, 66%, or 100% of the patient's data as the instance subset. Feature selection is conducted by $\theta_{min} = 0.1$. The vector contains the entire set of features as defined in Section 3.2.3. For features that are excluded due to the threshold $\theta_{min} = 0.1$, a value of zero is assigned.
- The size of the patient population that contributes to the Composite Multilinear Regression Model. In the first step of the training, the population begins with a random value and is subsequently updated to the agent's selection from the last step.
- The uniform value of θ_{min} applied across all individual models within the Composite Multilinear Regression Model. In the first step of the training, θ_{min} begins with a random value and is subsequently updated to the agent's selection from the last step.

Figure 5: Building Components of Environment's State Space

In practise, we have the option to make two distinct types of request to the environment. The first type of request involves asking for a new, randomly selected patient. During this query, the agent must specify two key parameters: the employed percentage over the total of patient's data and the Pearson Correlation Coefficient threshold value.

In response, the Environment provides the following information:

- The complete raw data for the selected patient
- A trained model specific to the selected patient, generated using the query parameters.
- Evaluation metrics, including the size of the evaluation dataset, the mean squared error, the mean absolute error, the root mean square error, and the coefficient of determination.

- Demographic details of the selected patient: Gender, Diagnosis, Age Group, and Weight Group.
- Statistics related to the model's training, including the number of features used and the number of instances in the training set.

The second type of request entails enquiring for the n most similar patients to a specific patient, denoted as S. The process of identifying the most similar patients to a specified individual leverages the following steps, utilising our pre-maked collection of models:

- Extract the demographic information (Gender, Diagnosis, Age Group, Weight Group) for S.
- Retain only patients who have a demographic profile identical to that of S.
- Calculate the Euclidean distance between the coefficients of S and those of the remaining patients, as described in Section 3.2.6.
- Select the n patients with the smallest calculated distances.

3.3.6 Deep Q-Learning Agent Design

The agent aims to optimise two key parameters: the size of the population within the Composite Multilinear Regression Model, consisting of patients similar to the one under examination, and a uniform Pearson Correlation Coefficient threshold that is applicable across all constituent models. Figure 6 provides an overview of the design decisions made during the construction of the agent.
The State Space is detailed in the previous Section 3.3.5. The length of the State Space vector consists of the 864 prediction features proposed in Section 3.2.3 employed as the coefficients taken from the single patient regression model for the patient under examination and the current population and the PCC threshold values for the Composite Multilinear Regression Model.

The Action Space encompasses a range of possible maneuvers the agent can execute, specifically targeting adjustments in both the population size and the Pearson Correlation Coefficient threshold. The actions are as follows:

- Value 0: Lower the PCC Threshold by a single increment,
- Value 1: Maintain the current PCC Threshold,
- Value 2: Increases the PCC Threshold by a single increment,
- Value 3: Reduce the population size by a single increment,
- Value 4: Sustain the current population size,
- Value 5: Augment the population size by a single increment.

Algorithm 2 Reinforcement Learning Agent Parameters and Q-Network Configuration

1: **Objective:** Optimize Population Size and PCC Threshold
2: **State Space:** Details in Section 3.3.5
Action Space:
- 0: Lower the PCC Threshold by one increment
- 1: Maintain the current PCC Threshold
- 2: Increase the PCC Threshold by one increment
- 3: Reduce the population size by one increment
- 4: Maintain the current population size
- 5: Augment the population size by one increment

3: **Valid Ranges of Population, Threshold:**
- Population: 1 to 10 (inclusive), in increments of 1
- Threshold: 0.1 to 0.9 (inclusive), in increments of 0.1

4: **Q-Network Configuration:**
- Input Layer: 866 units
- Hidden Layer: 100 units, Activation: ReLU
- Output Layer: 6 units (Actions)

5: **Learning Rate:** 0.001
6: **Loss Function:** Mean Squared Error (MSE)
7: **Optimizer:** Adam Optimizer [26]

8: **Termination Conditions:**
A: Agent maintains constant Population and PCC Threshold
B: Agent action produces out of range values for Population or PCC Threshold
C: Episode reaches maximum step

Figure 6: Reinforcement Learning Agent Parameters and Q-Network Configuration

The actions are represented in the Q-Network as an array consisting of six outputs. The output with the highest value dictates the agent's next action. An episode reaches its conclusion when the agent opts to maintain the current PCC (Pearson Correlation Coefficient) Threshold and, in the following step, chooses to sustain the existing Population size, or vice versa. The range of possible values for population sizes ranging from 1 to 10, inclusive, in increments of 1 and for PCC threshold values spanning from 0.1 to 0.9, also inclusive, in increments of 0.1.

The Q-Network consists of 1) an 866 units input layer, 2) a single hidden layer with 100 units and 3) an 6 units output layer. The activation function used is selected to be ReLU (Rectified Linear Unit). ReLU is a type of activation function that is widely used in deep learning models due to its simplicity. The function itself is quite simple, returning zero for all negative inputs and returning the input itself for all positive inputs $f(x) = \max(0, x)$[1].

For the learning rate, the value of 0.001 was selected for all experiments, considering it adequate to

avoid local minima. The loss function used in our model is the Mean Squared Error (MSE). Measures the average square difference between the estimated values and the actual values, providing a robust method for optimisation. We selected Mean Squared Error insteed of Mean Absolute Error to amplify the penalties on significant errors.

The Q-Network is optimised using the Adam optimizer, an adaptive learning rate optimization algorithm designed for training deep neural networks. It combines the advantages of two other extensions of stochastic gradient descent,AdaGrad and RMSProp, making it well-suited for handling sparse gradients and noisy data [26].

In future research, we plan to explore various neural network architectures and hyperparameters, aiming to compare their performance to further enhance our methodology.

3.3.7 Training Process

A summary of our training process is presented in Figure 7. As we can see, in each episode, the agent is introduced to a newly selected patient, chosen at random (making the batch size equal to 1). The patient is selected and introduced to the agent on the basis of a structured sequence. This sequence is designed to ensure that the agent is evenly exposed to all demographic subcategories. Specifically, for gender, a male patient is randomly chosen, followed by a female patient. Similarly, patients are sequentially selected from different diagnostic categories, Non-Diabetics, Type I Diabetics, and Type II Diabetics. The same structured approach is applied to age and weight groups, ensuring that agent performance benefits from a well-rounded learning experience across a diverse range of patient profiles.

Additionally, we employ a Replay Buffer with a capacity of 20, using a First-In, First-Out (FIFO) replacement strategy. The introduction of the Replay buffer serves in reducing the execution time (execution delays occur since each time a new patient is examined, his dataset must be preprocessed to make it suitable for the evaluation of the composite multilinear regression model).

For each episode's initial setup, the population and threshold values are set near their median values (5 for population and 0.5 for threshold). With a maximum value of 10 steps per episode, the agent is tasked with fine-tuning these parameters in an effort to achieve a Mean Absolute Error (MAE) for the Composite Multilinear Regression Model below 2.

An episode terminates if any of the following conditions are met:

- The agent's suggested an action that produce population or threshold beyond predefined limits (refer to Section 3.3.6).

Algorithm 3 Reinforcement Learning Episode Management and Training

1: **Initialization:**
- Initialize Replay Buffer with a capacity of 20.
- Set Discount Factor for future rewards to 0.95.

2: **For each episode:**
- Randomly select a new patient (Structured sequence for balanced demographic exposure)
- Initialize population=5 and threshold=0.5
- Set maximum steps per episode to 10

3: **while** Episode not terminated **do**
4: **Episode Termination Conditions:**
- Agent maintains both Population and Threshold constant in two consecutive steps
- Agent's action exceeds predefined limits for Population and Threshold
- Maximum number of steps per episode reached

5: Input State Space to Q-Network and recieve an action
6: Evaluate MAE for new Composite Multilinear Regression Model and calculate reward
7: Update Q-Network using Bellman Equation
8: Update ε value
9: **end while**
10: **Convergence Criteria:**
- Target Performance. Target MAE value lower than the Single-Patient Regression Model
- Performance encapsulated in Reward function. $MAE < 2$
- Optimal Performance. $MAE = 0$

Figure 7: Reinforcement Learning Episode Management and Training

- The agent chooses to keep both the Threshold and Population constant in two consecutive steps.
- The maximum number of steps per episode is reached.

Finally, in the Bellman Equation, the Discount Factor for future rewards is set to 0.95. During the training process, the Q-Network is updated after each new patient is examined. An increasing Cumulative Reward in each episode indicates that the agent is beginning to learn to optimise the values for population and threshold. The actual value of the Cumulative Reward is not meaningful by itself. The convergence criterion for the agent to reach optimal behaviour is an MAE value of zero. For our experimental work, the target is an MAE value lower than the equivalent value of the Single-Patient Regression Model.

4 Experiments and Evaluation

In the experimentation chapter, we empirically evaluate the methodology introduced in the previous chapter. Through a series of tests, we assess the extent to which the defined features can predict future insulin doses in various subsets of features and instances filtering scenarios, using the evaluation metrics of Section 3.2.7. We also provide a brief description of a toy example to showcase the basic steps of our method. Finally, the proposed Deep Q-Learning methodology is employed to train multiple agents and their performance is evaluated accordingly.

4.1 Development of the Experimentation Dataset

In the following section, we depict the characteristics of the empirical dataset used to conduct a series of experimental evaluations aimed at assessing the performance of the methodology proposed in the previous chapter. Our main data source is the Medical Information Mart for Intensive Care III (MIMIC-III) [23], a comprehensive database that contains a wide spectrum of information on various diagnoses, patient medications, and treatment procedures.

Given the specific focus of our study on insulin administration in both diabetic and non-diabetic patients, we extracted a relevant subset of data based on two distinct criteria, which we present below.

- We have selected patients who were admitted to the intensive care unit and during their stay, glucose measurements were taken. Additionally, at least one type of insulin was administered to these individuals during their stay.
- To avoid possible bias in our study, we have opted to include only the initial admission to the intensive care unit for each patient in our analysis.

The resulting set of patients comprises 8,225 individuals, each corresponding to a unique admission to the intensive care unit. For each of these individuals, we extracted a set of features, which describe them demographycally, their diagnosis, and how their glucose levels react to insulin administration. Briefly stated, we extracted information from MIMIC-III tables *Admissions*, *Icustays*, *Patients*, *Chartevents*, *Inputsevents_mv* and *Diagnoses_ICD*. Complementary tables *D_ICD_Diagnoses* and *D_Items* were also used to correspond diagnoses, measurements, and drug codes to equivalent string labels.

For presentation purposes, we divide the extracted data into five distinct categories.

- Demographic attributes of the patients: *Gender*, *Age*, *Death Location*, and *Weight*.
- Time related to the admission of the patient to the intensive care unit.
- Information on diagnosis, specifically the classification of patients as either *Non-Diabetic*, *Diabetic Type I* or *Diabetic Type II*.

- Details related to *Insulin Administration*, including both *Time of Administration* and *Dosage*.
- Information related to *glucose measurements*, specifically, the *measurement time* and the respective *glucose levels*.

4.1.1 Dataset Population Demographic Profile

The demografic description of our selected dataset serves the purpose of providing general description of our selected individual pinpointing any potential bias.

Gender and *Age* information were extracted from columns *gender*, *dob* in MIMIC table *Patients*. The attribute *Death Location* takes 3 unique values "Alive", "Death in ICU" and "Death after ICU". Death-related information was taken from columns *dod* and *expire_flag* of MIMIC table *Patients* in combination with ICU stay times information from the MIMIC table *ICUStays*. Values of *Weight* attribute was sollely extraxcted from column *patientweight* of table *Inputevents_mv*.

In Table 1, it is observed that males make up the majority 56.3% of our data set. Furthermore, 7.7% of the patients died during their stay in the ICU. Additionally, the data set comprises a substantial number of Non-Diabetic and Type II Diabetic individuals, with Type I Diabetics representing a smaller subset.

Categorical Features	Percentages
Gender	59.4% male, 40.57% female
Diagnosis	56.3% Non Diabetic, 3.2% Diabetic Type I, 40.4% Diabetic Type II
Death Location	71.9% alive, 7.7% death in ICU, 20.2% death after ICU

Table 1: Percentage Distribution of *Gender*, *Diagnosis*, and *Death Location* in the 8,225 Population

Table 2 presents the central tendency metrics of the mean and median values and the dispersion metric of the interquartile range (IQR) for the numeric variables of patient's *Age* and *Weight*. We observe that a substantial proportion of our patient population falls into the middle- to old-age categories.

Numeric Features	Mean	Median	IQR
Age	65.5	67.0	21.0
Weight	84.5	81.0	27.9

Table 2: Central Tendency and Dispersion of *Age* and *Weight*

Another potentially useful feature is whether the patient has passed away during the current stay in ICU or not. The distribution of deaths in ICU is shown in Table 1, while the distribution of *mortality in the ICU* in relation to *Age* and *Weight* is illustrated in Figure 8. For simplicity reasons, we will not include this feature in our experiments. In future work, we may investigate whether patients who

died in the ICU differ in terms of insulin dosing and glucose measurements compared to the rest of the population.

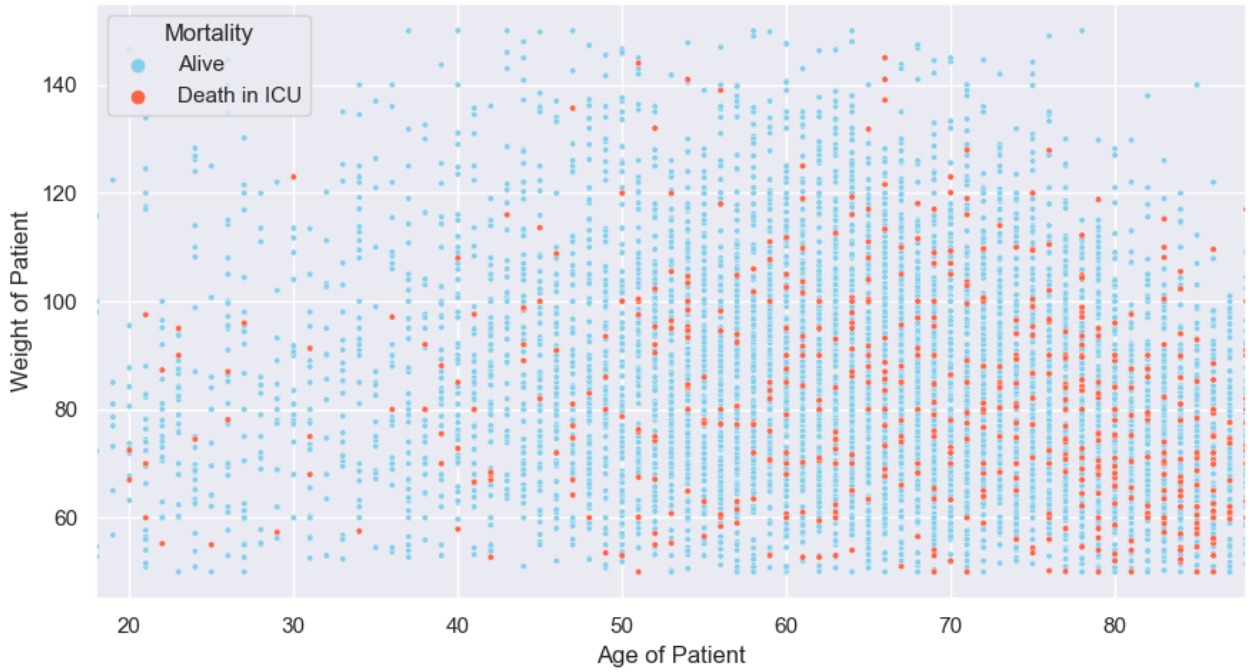

Figure 8: Distribution of *Mortality in the ICU* in relation to Patient's *Age* and *Weight*

4.1.2 Extraction of Time Related Information

Time-related information plays a critical role in our produced dataset, giving the order of the sequence of caregivers' observations (Glucose Measurements) and actions (Insulin Administrations).

There exist three different categories of time information in which we are interested:

- The time the patient enters the ICU (column *intime* in table *ICUStays*)
- The time of each glucose measurement (column *charttime* in table *Chartevents*)
- The starting and end time of each Insulin administration (column *starttime* and *endtime* in table *Chartevents*)

Using the above timestamps, we convert each glucose measurement and insulin administration timestamp as a triple of: a) the time elapsed of the event since ICU entry, b) day counter in ICU, and c) hour of the day. Using this approach, we can order the sequence of events for each patient, but we can also group the events by specific hours of the day.

For example, if a patient enters the ICU at 16:00 08/05/2023 and a glucose measurement was taken at 18:00 on the same day, then the time information is converted as: 7200 seconds, 1st day, 18 hours of

day. In the case of insulin administration timestamps, we follow the same approach, but we generate a series of time triplets, one for each minute insulin was administered. For example, if a patient enters the ICU at 16:00 on 08/05/2023 and 3 units of insulin were administered between 18:00 and 18:03 on the same day, then the time information is converted to (7200, 1, 18), (7260, 1, 18), and (7120, 1, 18). Note that 1 Unit of Insulin corresponds to each timestamp.

Another important parameter to consider is the duration of stay in the ICU, as shown in Figure 9. Most of the patients had stays that were less than a week in duration. In subsequent studies, it would be worth exploring whether the length of stay introduces bias, especially given the potential for repetitive patterns in administrations during extended stays.

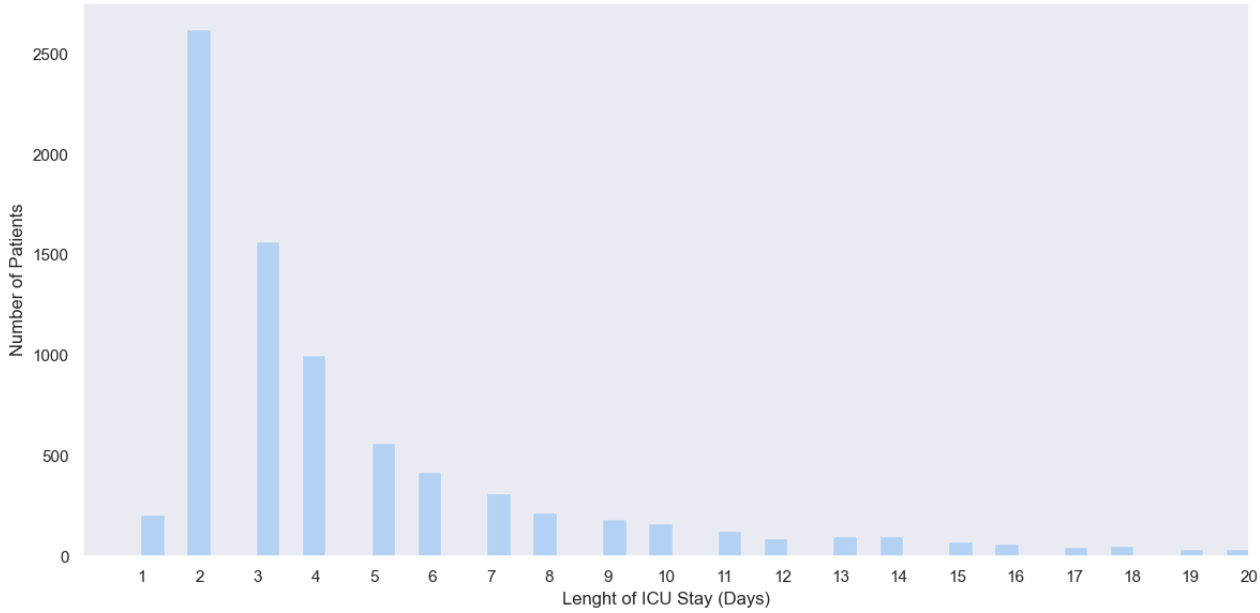

Figure 9: Number of Patients per length of ICU Stay in days

4.1.3 Extraction of Diabetes Diagnosis Information

MIMIC table *Diagnoses_ICD* contains a list of diagnoses per patient ordered by severity (column *seq_num*). We devide the examining patients in 3 categories: *Non Diabetic*, *Diabetic Type I*, *Diabetic Type II*. Table 3 illustrates the populations for each category of patients.

Using column *icd9_code* we tag rows based on their value prefix. Values with prefix 250 define diabetes [36]. To further split diabetic patients into Type I and Type II, we employ column *long_title* of table *D_ICD_Diagnoses* and devide patients based on the substring *TYPE I* and *TYPE II*. Note that we define a patient as diabetic if the prefix 250 exists, regardless of severity order. Additionally, Appendix Table 16 provides a detailed view of all diabetes subcategories in our dataset. In future

work, the experimentation dataset can be filtered based on specific subcategories.

Diagnosis	Number of Patients
Non Diabetic	4526
Diabetic Type-I	263
Diabetic Type-II	3323

Table 3: Number of *Non Diabetic*, *Diabetic Type-I* and *Diabetic Type-II* Patients

4.1.4 Extraction of Insulin Administrations Amounts

Information about insulin administration is stored in MIMIC III tables *Inputevents_CV* and *Inputevents_MV*. Due to the incompatibilities between these two tables and the more accurate time information of table *Inputevents_MV* only the latter will be used in our experiments. The data of the 8,225 individuals mentioned in Section 4.1 was extracted exclusively from table *Inputevents_MV*.

According to Johnson et al. [22] data in table *Inputevents_MV* were organised using the Philips iMDSoft Metavision System, a workflow and data management system specifically designed for use in critical care environments. The system was used at Beth Israel Deaconess Medical Centre between 2008 and 2012.

During the first step of the extraction process, the MIMIC table *D_Items* was filtered to keep rows that contain the value *inputevents_mv* in column *param_type*. These rows were further filtered to rows where column *label* contains substring *Insulin*. Finally, using the final list of rows, a dictionary was constructed that matches the values between columns *itemid* and *label*. Using this dictionary, we can replace values of column *ItemID* of the table *Inputevents_MV* with the label of the substance (for example, Insulin - Glargine).

As we can see in Table , *Insulin - Humalog 75/25*, *Insulin - NPH* and *Insulin - 70/30* are rarely administered. For simplification purposes *Insulin - Humalog 75/25* was replaced with *Insulin - Humalog* and *Insulin - NPH* and *Insulin - 70/30* was replaced with *Insulin - Glargine*.

In Philips iMDSoft Metavision System, the amount of insulin administered is recorded as the total number of units between $t_{starttime}$ and $t_{endtime}$ (column *starttime* and *endtime* in table *inputevents_mv* respectivly) or as rate of units per hour (column *rate* in table *inputevents_mv*). For the purpose of greater flexibility during the generation of the features, we convert all records into units of insulin per minute.

		Insulin Action Properties [2], [3], [4], [7], [17]			
Insulin Label	Occurrences	Insulin Type	Onset	Peaks	Duration
Insulin - Regular	89496	Short Acting	30 Min	2-3 Hours	6-8 Hours
Insulin - Humalog	18028	Rapid Acting	15 Min	1 Hours	2-4 Hours
Insulin - Glargine	5697	Long Acting	1-1.5 Hours	Peakless	24 Hours
Insulin - NPH	2139	Intermediate Acting	1-3 Hours	6-8 Hours	12-16 Hours
Insulin - 70/30	311	Short Acting	30 Min	2-3 Hours	6-8 Hours
Insulin - Humalog 75/25	86	Rapid Acting	15 Min	1 Hour	2-4 Hours

Table 4: Number of Insulin Administrationsin the experimentation dataset and Insulin Action Properties per Insulin Type

4.1.5 Extraction of Glucose Levels Measurements

Information about glucose measurements for 8,225 patients was extracted from MIMIC table *Chartevents*. Similarly to insulin administrations, Table *D_Items* was used to create a dictoanary to match *itemid* codes to measurement labels (where column *param_type* in table *D_Items* takes value *chartevents*). We filtered the rows of *chartevents* and keep only those which column *itemid* coresponts to a label that contains the substring *glucose*. Table 5 presents the number of occurrences per label.

Glucose Label	Occurrences
Glucose Fingerstick	151089
Glucose (Serum)	65315
Glucose (Whole Blood)	46629
Glucose	2090
Glucose (70-105)	2080

Table 5: Glucose Measurements in the Experimentation Dataset

All labels in Table 5 represent Whole Blood Glucose except label *Glucose (Serum)*. Serum/Plasma has a higher water content than whole blood (which also contains red and white blood cells, platelets, etc.). Therefore, the same amount of glucose is present in a smaller volume in serum/plasma, leading to higher measured concentrations. The World Health Organisation (WHO) has devised a conversion factor of 1.12 [27]. Calculating the average value using all whole blood glucose measurements (column *valuenum* in table *Chartevents*) in our data set and doing the same with serum glucose measurements, we calculated a factor of 1.1306 (which is approximately equal to the WHO number). Using this number, we converted all of our measurements to total blood glucose. In our data set, glucose is universally measured in mg/dL (column *valueom* in table *Chartevents*).

4.1.6 Definition of Dependent and Independent Variables

Following the methodology of Section 3.2.3, we generate a set of Dependent and Independent Variables. Depended variables span durations of last 15 minutes to 36 hours incrementing in steps of 15 minutes, while independent variables range of future 15 minutes to 2 hours in steps of 15 minutes.

Table 6 presents the Mean, Standard Deviation and Interquartile Range (IQR) values for the glucose level and insulin amounts for the shortest period of our independent variables. It is observed that long-acting insulin is administered in the largest amounts in comparison to short- and rapid-acting insulin.

The Standard Deviation, often denoted by σ, is a measure of the dispersion or spread of a set of values. It is defined as the square root of the variance, $Var(X)$, of a random variable X. Mathematically, the standard deviation is expressed as: $\sigma = \sqrt{Var(X)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$ where N is the number of observations, x_i represents each individual observation, and μ is the mean of all observations. The Interquartile Range (IQR) is a measure to describe the spread of data and is defined as $\text{IQR} = Q_3 - Q_1$ where Q_1, is the value below which 25% of the data falls, while the third quartile, Q_3, is the value below which 75% of the data falls.

Feature Description	Mean	σ	IQR
Average Glucose Level (mg/dL)	155.22	61.6	60.7
Total Amount of Insulin (Units)	2.56	6.02	1.67
Rapid-Acting Insulin (Units)	5.17	4.68	4.0
Short-Acting Insulin (Units)	1.82	2.82	1.5
Long-Acting Insulin (Units)	25.39	20.16	25.0

Table 6: *Mean Value*, *Standard Deviation* and *Interquartile Range* for Independent Variables last 15 minutes

For a more detailed insight into 24-hour insulin administrations, refer to Appendix Table 17, which displays statistics for intervals of 3, 6, 12, and 24 hours. When viewed alongside 6, it is evident that even though long-acting insulin is given in higher doses, short-acting insulin sees more frequent administration.

To set a reference for subsequent evaluations in the following sections, we calculate the mean value, standard deviation, and interquartile range for glucose levels and insulin amounts over 1 and 2-hour periods. The findings are detailed in Table 7.

Numeric Features	Mean	σ	IQR
Average Glucose Level next 1 Hour (mg/dL)	151.65	61.25	59.25
Average Glucose Level next 2 Hours (mg/dL)	150.54	59.29	57.88
Total Insulin next 1 Hour (Units)	6.97	11.17	5.34
Total Insulin next 2 Hours (Units)	12.27	18.16	9.88
Rapid Acting Insulin next 1 Hour (Units)	5.66	5.38	4.5
Rapid Acting Insulin next 2 Hours (Units)	6.22	6.07	6.0
Short Acting Insulin next 1 Hour (Units)	5.83	8.25	4.71
Short Acting Insulin next 2 Hours (Units)	10.75	15.19	8.08
Long Acting Insulin next 1 Hour (Units)	27.12	21.93	30.0
Long Acting Insulin next 2 Hours (Units)	28.08	22.63	30.0

Table 7: *Mean Value*, *Standard Deviation* and *Interquartile Range* for Independent Variables last 3, 6, 12 and 24 hours

4.2 Toy Example on Patient 50315

In a hospital setting, caregivers can get a sense of a patient's response to varying insulin doses at different times of the day by adjusting subsequent administrations based on a ledger of the patient's preexisting data. Each row of the table contains information on insulin intake and glucose levels for one day, with subsequent rows detailing the data for the following days. The information for each day is focused predominantly on the meals of the patients. This caregiver activity is also reflected in our experimentation datasets, where the periods before and after food intake are primarily targeted. Figure 10, derived from our experimentation dataset, depicts the number of patients with at least one glucose measurement and insulin administration for each hour of the day. Intervals before and after meals show increased activity.

In this section, a simplified version of our methodology is presented using real data from our dataset of a 68-year-old woman with a weight of 62 kilogrammes and no diagnoses of diabetes. During his first stay in the ICU, which lasted 18 hours, she received 54.5 units of regular insulin, which is higher than the *0.55 X Total Weight in Kilogrammes* general rule. The average glucose level was 123 mg/dL with a range from 96 to 182 mg/dL. Table 8 illustrates the patient's blood glucose-insulin ledger. Since our datasest uses insulin infusion rates per minute we produced approximate absolute values using as meal times 10:00, 13:00, and 17:00.

The summarised data of Table 8 was captured in 72 time itervals in our dataset, with a duration of 15 minutes each. Appendix Figure 22 illustrates the administration of Short-Acting Insulin to the patient, as this was captured by the target variable Next-1-hour Short-Acting Insulin Dose.

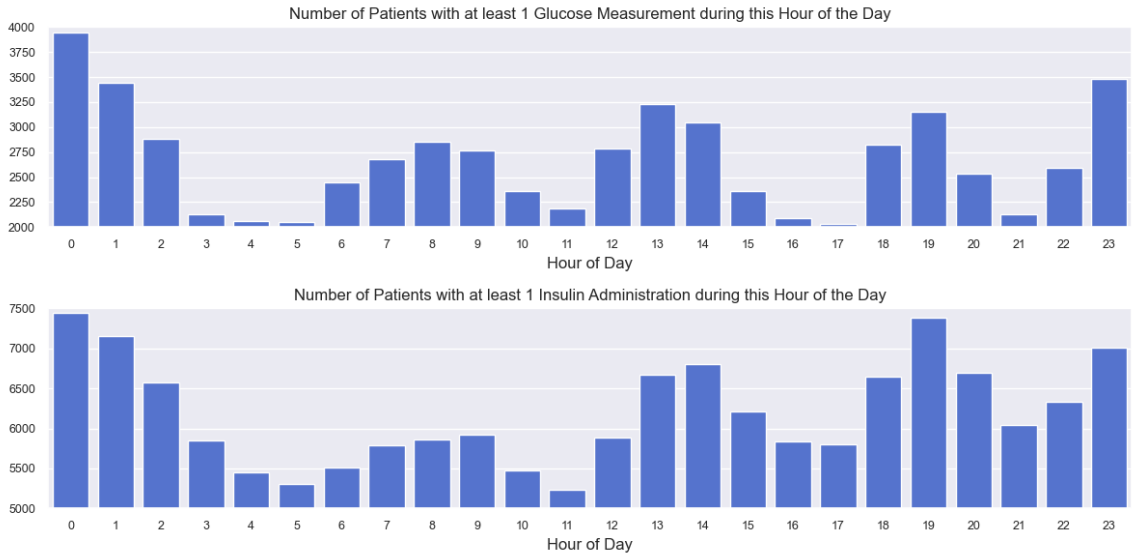

Figure 10: Glucose Measurements and Insulin Administrations per Hour of the Day for all ICU Patients

	Breakfast			Lunch			Dinner			Night		
	Before	Insulin	After	Before	Insulin	After	Before	Insulin	After	Before	Insulin	After
Day 1.	-	-	164	98	7.5	109	120	6.7	132	114	32	109
Day 2.	182	8	-	-	-	-	-	-	-	-	-	-

Table 8: Blood Glucose-Insulin Ledger of Patient 50315 during an 18-hours ICU Stay

An alternative approach to viewing patient data is to plot the amounts of insulin doses on glucose measurements. In this way, we can pinpoint any possible relations between them. Figure 11 illustrates the dispersion between the Past-2-Hours Short-Acting insulin administered doses over the Next-2-Hours average Glucose Measurements. A linear relationship is evident.

In the first stage of this illustrative example, we partition the patient's data into two subsets: the initial 50% for training and the remaining 50% for evaluation. Using all predictive attributes from Section 3.2.3, we train a multilinear regression model to forecast the short-acting insulin doses for the next hour. Long features periods with missing data were excluded. Using the test subset, we obtained an MAE of 1.41 and an RMSE of 3.54. In this preliminary example, due to the limited data, we opted not to perform any feature selection. Appendix Figure 23 illustrates the real and predicted next-1-hour short-acting insulin doses over time.

In the next stage, we performed the same experiment, training a multilinear regression model to forecast the short-acting insulin doses using data from 200 randomly selected patients, generating a training dataset of 2,530 instances. To maintain consistent settings, we also chose not to perform

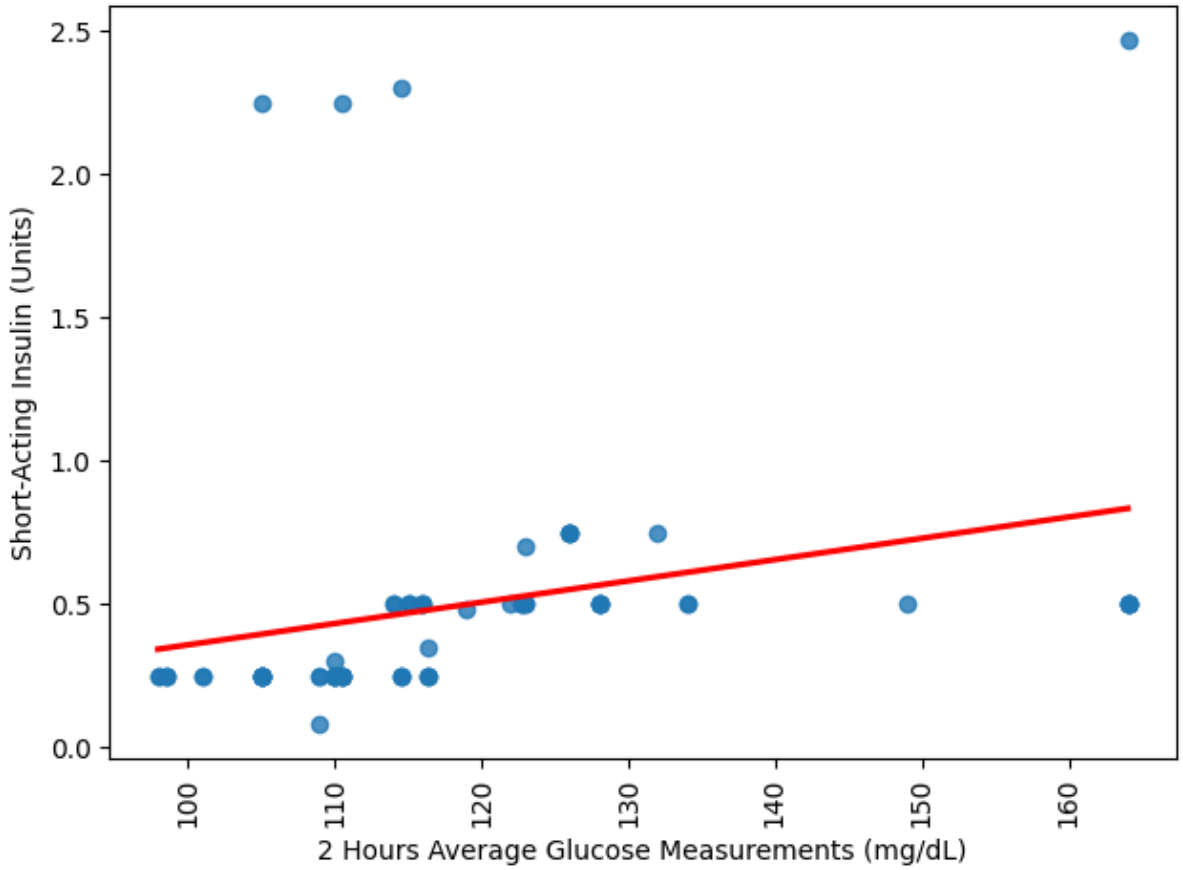

Figure 11: Example Patient's Past-2-Hours Short-Acting insulin administered doses over Next-2-Hours average Glucose Level

feature selection in this instance. Using the test subset from the previous setting (the latter half of patient 50315's data), we recorded an MAE of 3.19 and an RMSE of 11.49. Figure 12 illustrates the actual and predicted Next-1-Hour Short-Acting Insulin doses over time using multiple patients in the training dataset. Note that despite inferior performance, the second approach provided predictions for the duration of the patient's stay in the ICU.

The focus of our next experimental work can be outlined as follows: To predict a patient's insulin dose at a specific time t, what types and how many patients should be chosen to train a multilinear regression model in order to outperform a similar model trained on an existing subset of patient data/ This investigation also extends to the selection of relevant features in the training dataset from these patients.

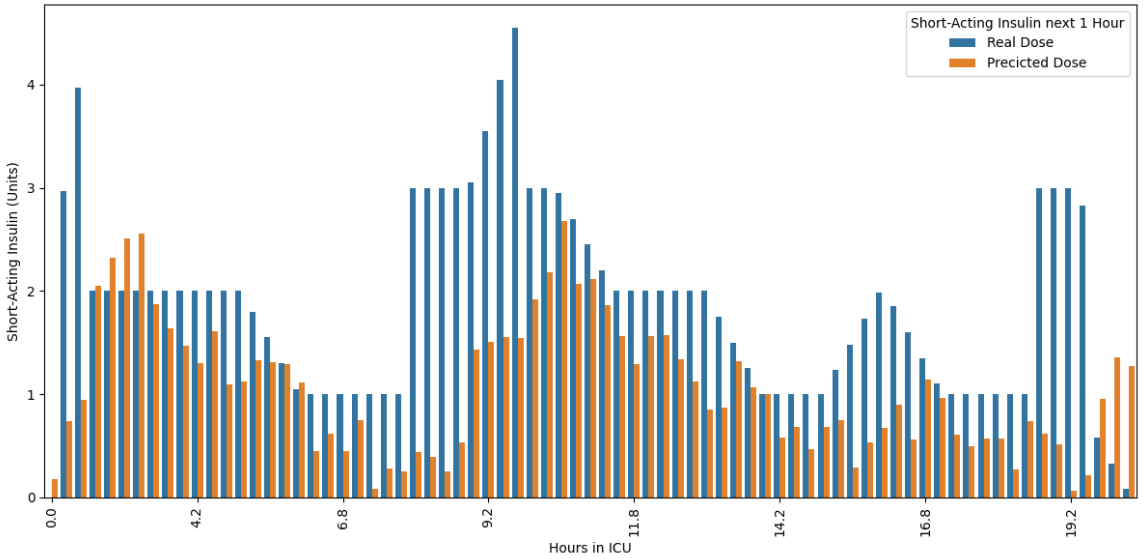

Figure 12: Real and Predicted Next-1-Hour Short-Acting Insulin Doses using Multilinear Regression Model trained on 200 randomly-selected patients data over Time

4.3 Learning Curve Analysis

In this section, we evaluate the performance of Multilinear Regression in predicting future insulin doses and average glucose measurements using our experimentation dataset. This evaluation utilizes the set of features outlined in Section 4.1.6 and follows the methodology detailed in Chapter 3. Specifically, we perform a learning curve analysis with increasing amounts of training data by increasing the number of patients in the training dataset. The objective of this section is to determine if larger training populations perform better than smaller ones, a conclusion that will aid us in designing our reinforcement learning experiments more efficiently.

We perform two classes of experiments. In the first, we performed the experiments with randomly selected patients, while in the second, selection is based on identical demographic groups. The minimum size for the patient population in both scenarios is set at 50 individuals, as smaller sample sizes do not produce accurate results.

4.3.1 Experiments on Randomly Selected Patients

In this subsection, we focus on patients selected purely at random. By utilising a random sampling approach, we aim to provide a baseline understanding of the model's responsiveness to data quantity without introducing potential biases from structured subsets. The results offer an unfiltered perspective on how performance trends evolve with the incremental addition of random data points.

In our first baseline experiment, we trained 200 separate multilinear regression models and then averaged their results. In each model, we selected a sample of 50 patients from the overall group of

8,225 individuals. Additionally, 20 randomly selected patients were chosen to evaluate each model. It is important to note that the patients in this experiment are distinct from those mentioned in the preprocessing section (refer to Section 3.2.5). Furthermore, based on the set of training attributes presented in Section 3.2.3, we applied the feature scoring methodology outlined in Section 3.2.4 using Pearson's correlation coefficient (PCC). We retained features with a PCC exceeding 0.1, effectively filtering out completely irrelevant attributes.

Table 9 presents the results for the tasks of predicting the future short-acting insulin dose and mean glucose level over 1 and 2 hours highlighting the best-performing metrics with bold text. In all cases, we observe low R2 and high MAE. However, MAE is less than the standard deviation value. When assessing individual training outcomes, we observed instances where certain training sessions yielded entirely irrelevant results. Although these occurrences were infrequent, their presence further underscores concerns about the randomly selected training datasets.

	Dataset			Actual Values				Prediction Values				Prediction		
Target	Train. Sz	Test Sz	Features	Min	Max	Mean	σ	Min	Max	Mean	σ	MAE	RMSE	R2
Average Glucose														
Next 1 hour	4106	1407	583	43.3	477.0	143.2	50.9	0.0	4149.5	146.7	119.0	33.3	41.8	0.15
Next 2 hours	4974	2054	564	56.5	430.5	141.2	47.3	0.0	1147.2	145.6	55.1	**29.5**	40.7	0.25
Short-Acting Insulin														
Next 1 hour	4492	4492	521	0.1	51.6	5.9	5.6	0.0	62.8	5.9	6.4	**3.2**	3.9	0.44
Next 2 hours	3772	911	572	0.0	94.8	12.2	11.4	0.0	36.8	7.7	6.0	5.6	8.9	0.31

Table 9: Average Performance of Multilinear Regression in 200 Tests: 1 and 2 hours Predictions based on 50 randomly-selected training patients.

Applying the same methodology described above to a dataset of randomly selected training patients with a population of fewer than 50 yielded significantly worse results. For this reason, we have chosen not to include those results.

To evaluate the findings of Table 9 against training datasets of varying sizes, we focused on the better performing tasks of predicting future short-acting insulin doses and forecasting average glucose levels over the next 1 hour.

We carry out the same methodology as presented above in a data set of randomly selected training patients with population between 50 to 100 patients. For each training population, we train 5 separate multilinear regression models and average the results. We kept all the parameters identical to the baseline experiment above and only altered the training data set size.

Table 10 presents the results of the evaluation. For the task of predicting the glucose level, MAE dropped from 33.3 to 32.3, which indicated a very small increase in performance. For the task of

predicting the netx-1-hour short-acting insulin, MAE reduced from 3.2 to 3 and R2 increased from 0.44 to 0.54, presenting a slightly better performance.

	Dataset			Actual Values				Prediction Values				Evaluation		
Target	Train. Sz	Test Sz	Features	Min	Max	Mean	σ	Min	Max	Mean	σ	MAE	RMSE	R2
Average Glucose														
Next 1 Hour	5359	1184	597	39.0	580.6	140.7	47.4	0.0	764.2	158.7	70.5	**32.3**	42.4	0.21
Short-Acting Insulin														
Next 1 Hour	6520	1179	609	0.1	54.3	6.8	6.6	0.0	47.3	7.5	7.6	**3.0**	4.0	0.54

Table 10: Average Performance of Multilinear Regression in 1 Hour Predictions based on randomly-selected training patients dataset with populations from 50 to 100.

Figure 13 provides a clearer illustration of how Multilinear Regression performs in predicting the Next-1-Hour Short-Acting insulin dose as the number of patients in the training data set increases from 50 to 100.

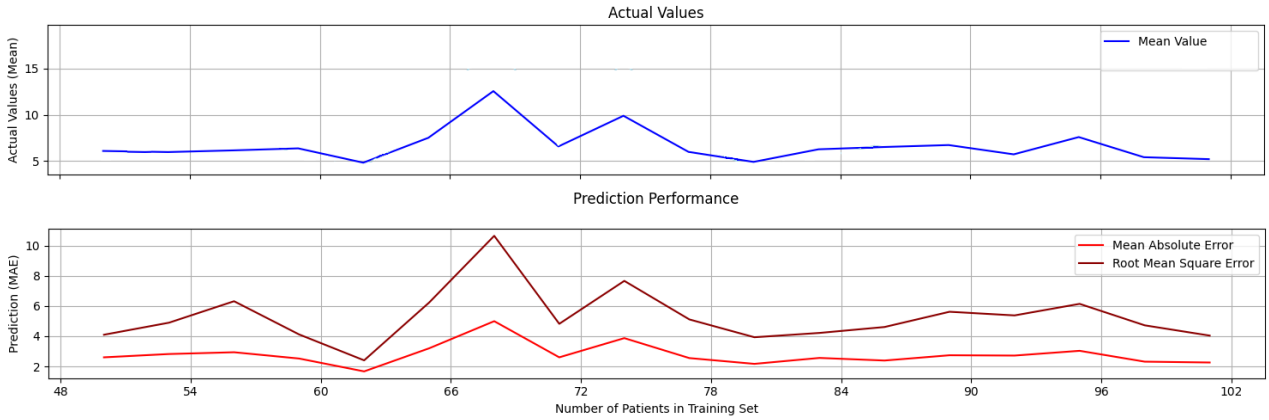

Figure 13: Average Performance of Multilinear Regression in 1 hour Short-Acting Insulin predictions based on randomly-selected training patients dataset with population intervals of 5 from 50 to 100.

Similarly, Figure 14 presents the results for the task of the Next-1-Hour Average Glucose Level as the number of patients in the training dataset increases from 50 to 100.

Finally, we investigate the performance in larger datasets following the same methodology with population intervals of 50 from 100 to 1000. On average, this resulted in a training dataset size of 10.000 data points. For both tasks of short-acting insulin dose and forecasting average glucose levels, performance was not affected. Figure 15 illustrates the individual training results for the short-acting insulin dose task.

4.3.2 Experiments on Patients in Identical Demographic Groups

In the preceding section, we performed a learning curve analysis on randomly selected data sets, offering insights into how our models perform at a baseline level. By targeting to pinpoint performance in-

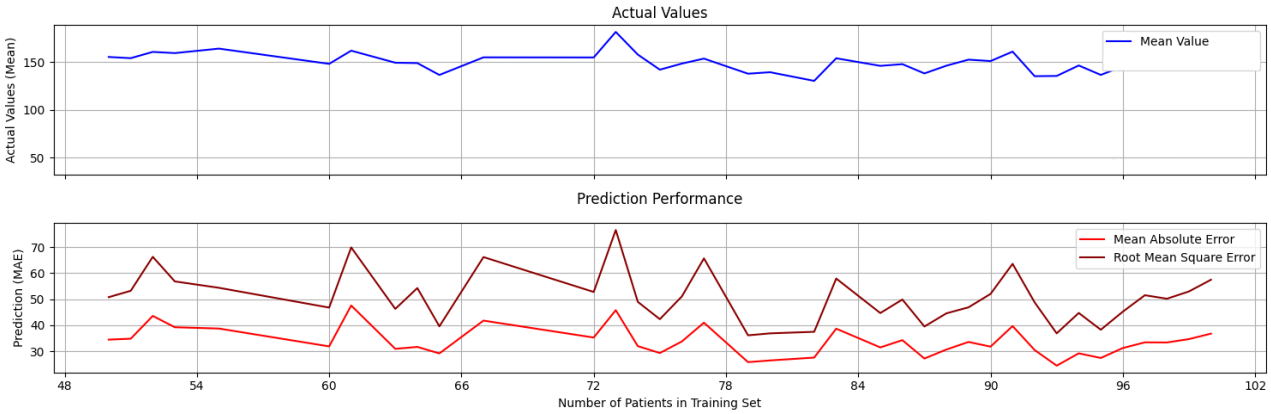

Figure 14: Average Performance of Multilinear Regression in 1 hour average glucose levels predictions based on randomly-selected training patients dataset with population intervals of 5 from 50 to 100.

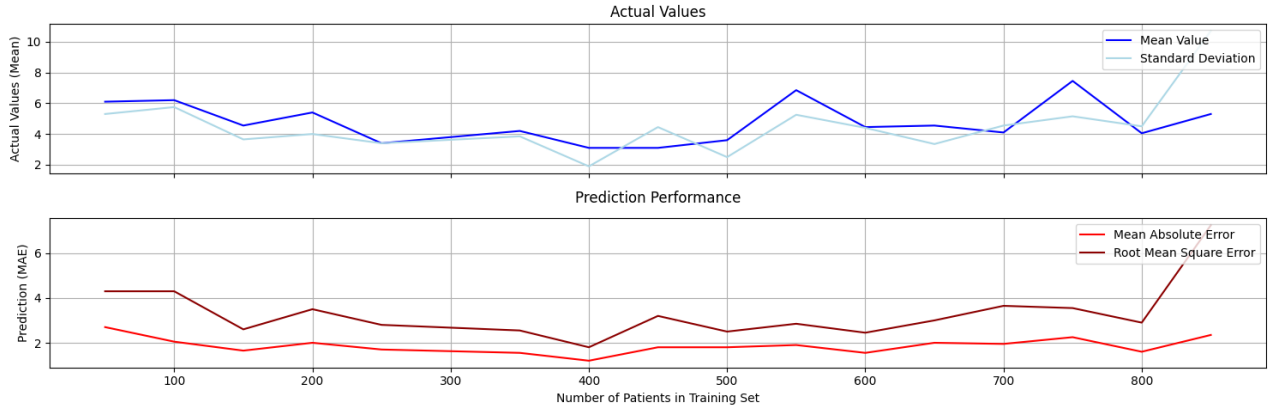

Figure 15: Average Performance of Multilinear Regression in 1 hour Short-Acting Insulin predictions based on randomly-selected training patients dataset with population intervals of 50 from 100 to 1000.

creases, we focus on examining how these learning curves behave in demographically similar datasets.

Following the guidelines of the clustering process presented in Section 3.2.6 using our experimetation dataset of 8,225 individuals we extract subsets of the following demographic categories:

- Based on *Gender* we cluster 4885 *Male* and 3336 *Female* patients.
- Based on *Weight Group* we cluster 2853 patients of *Low Weight* (less than 75 kg), 3438 patients of *Middle Weight* (from 76 to 100 kg), 1092 patients of *High Weight* (from 100 to 120 kg) and 409 patients of *Extreme-High Weight* (more than 120 kg).
- Based on *Diagnosis* we cluster 263 *Diabetic Type I* patients, 3323 *Diabetic Type II* patients and 4526 *Non Diabetic* patients.
- Based on *Age Group* we cluster 850 *Young* patients (less than 45 years), 3001 *Middle-Age* patients (from 46 to 65 years) and 4374 *Old* patients (more than 66 years).

Using the same setup as our 50 patients baseline experiment, we trained 200 separate multilinear regression models and averaged the results. Each model was trained on 50 and evaluated on another 20 patients. This time, patients from the same demographic group were chosen during training and evaluation, and we recorded the differences in performance in each group. Table 11 shows the evaluation results highlighting the best-performing metrics with bold text.

Group		Dataset			Actual Values		Prediction Values		Evaluation (with demogr.)		Gain	
Name	Value	Train. Sz	Test Sz	Features	Mean	σ	Mean	σ	MAE	R2	MAE	R2
Average Glucose next 1 Hour												
Gender	Female	4868	1236	177	149.9	17.8	142.1	9.3	13.0	0.16	**20.3**	-0.04
Gender	Male	3458	1245	176	149.9	16.8	149.4	9.5	12.9	0.18	**20.4**	-0.04
Diagnosis	Non Diabetic	4868	1247	169	151.9	18.8	149.4	9.3	13.0	0.18	**20.3**	-0.04
Diagnosis	Diabetic Type-I	4453	1346	160	150.0	17.8	147.4	9.7	13.2	0.14	**20.1**	-0.14
Diagnosis	Diabetic Type-II	2348	1245	179	149.1	18.8	148.5	9.4	13.1	0.16	**20.2**	-0.04
Age Group	Young	5468	1215	164	148.4	16.8	145.9	9.8	13.2	0.14	**20.1**	-0.14
Age Group	Middle Age	4848	1212	173	152.9	17.8	150.7	9.4	13.1	0.16	**20.2**	-0.04
Age Group	Old	4868	1116	176	149.9	17.3	147.3	9.7	13.0	0.17	**20.3**	-0.04
Weight Group	Low	4868	1211	176	149.9	17.2	145.9	9.5	13.2	0.15	**20.1**	-0.04
Weight Group	Middle	4802	1246	176	150.2	17.4	148.5	9.4	13.0	0.17	**20.3**	-0.04
Weight Group	High	2368	1214	164	149.9	16.8	151.5	9.6	13.0	0.17	**20.3**	-0.04
Weight Group	Extreme-High	1868	1216	155	149.8	17.3	145.2	9.6	13.2	0.15	**20.1**	-0.04
Short-Acting Insulin next 1 Hour												
Gender	Female	2676	498	530	5.4	3.2	6.2	5.4	2.4	0.59	**0.8**	**0.16**
Gender	Male	2554	592	539	5.5	4.9	5.9	5.1	2.2	0.58	**1.0**	**0.16**
Diagnosis	Non Diabetic	2676	498	527	5.4	6.9	6.4	3.7	2.5	0.58	**0.7**	**0.16**
Diagnosis	Diabetic Type-I	2554	592	537	5.0	5.0	5.4	3.2	2.2	0.61	**1.0**	**0.16**
Diagnosis	Diabetic Type-II	2554	592	540	5.2	4.0	5.2	4.5	2.6	0.61	**0.6**	**0.16**
Age Group	Young	2554	592	541	5.1	4.1	5.3	4.7	2.3	0.61	**0.9**	**0.16**
Age Group	Middle Age	2676	498	530	4.5	5.0	4.3	9.6	2.8	0.58	**0.4**	**0.16**
Age Group	Old	2676	498	531	5.4	4.0	5.4	3.4	2.3	0.59	**0.9**	**0.16**
Weight Group	Low	2390	444	511	5.0	3.4	6.5	3.8	2.5	0.46	**0.7**	**0.06**
Weight Group	Middle	2676	498	532	5.1	3.9	5.5	4.8	2.6	0.58	**0.6**	**0.15**
Weight Group	High	2676	498	532	5.5	5.3	6.3	4.3	2.3	0.49	**0.9**	**0.06**
Weight Group	Extreme-High	2676	498	534	5.9	3.0	5.4	3.8	2.1	0.56	**1.1**	**0.17**

Table 11: Average Performance of Multilinear Regression Over 200 Tests: 1-Hour Predictions Based on Training and Evaluation Datasets from 50 Demographically Identical Patients.

The last two columns, *MAE* and *R2*, present the gains achieved in the individual group compared to the equivalent results of the baseline setting with 50 randomly selected patients. It is observed that there is a gain in all cases.

In the last experiment, we investigated the performance of multilinear regression on larger training datasets. Specifically, we perform the same tests as above on demographically identical training and test patient datasets, with population intervals of 50 from 100 to 1000. For each population size, we train 5 models and average the results. The test population consists of 20 patients, randomly selected from the same demographic group. In cases where the desired population size exceeds the maximum number of available patients in that specific group (e.g., there are 409 available patients in

the 'Extreme-High Weight' group and 263 in the 'Diabetic Type I' group), we include all available patients in the training population, excluding those randomly selected for testing purposes.

Group		Dataset			Actual Values		Prediction Values		Evaluation (using demogr.)		Gain	
Name	Value	Train. Sz	Test Sz	Features	Mean	σ	Mean	σ	MAE	R2	MAE	R2
Short-Acting Insulin next 1 Hour												
Gender	Female	13632	764	511	7.7	7.4	7.6	8.3	2.1	0.65	0.5	0.25
Gender	Male	13631	794	510	7.6	7.3	7.4	8.1	2.1	0.64	0.5	0.15
Diagnosis	Non Diabetic	13421	781	515	7.6	7.3	7.2	7.3	2.1	0.64	0.5	0.15
Diagnosis	Diabetic Type-I	13631	794	515	7.6	7.3	7.4	8.4	2.1	0.64	0.5	0.15
Diagnosis	Diabetic Type-II	13632	764	509	7.7	7.4	7.3	7.2	2.1	0.65	0.5	0.25
Age Group	Young	13631	794	515	7.6	7.3	7.4	7.4	2.1	0.64	0.5	0.15
Age Group	Middle Age	13631	794	512	7.6	7.3	7.2	7.3	2.1	0.64	0.5	0.15
Age Group	Old	13631	794	510	7.6	7.3	7.3	7.2	2.1	0.64	0.5	0.15
Weight Group	Low	13631	794	512	7.6	7.3	7.2	7.5	2.1	0.64	0.5	0.15
Weight Group	Middle	13631	794	511	7.6	7.3	7.3	8.2	2.1	0.64	0.5	0.15
Weight Group	High	13631	794	514	7.6	7.3	7.4	7.2	2.1	0.64	0.5	0.15
Weight Group	Extreme-High	13421	781	520	7.6	7.3	7.2	8.4	2.1	0.65	0.5	0.15

Table 12: Average Performance of Multilinear Regression in 1 hour Short-Acting Insulin predictions based on demographically identical training patients dataset with population intervals of 50 from 100 to 1000.

4.3.3 Conclusions

Repetition of the above experiments multiple times leads to the conclusion that demographically identical patients in the training data set produce better results than randomly selected patients. Furthermore, there exists variability between the results of individual experiments, which leads us to consider not using randomly selected patients (either from the general population or from the same demographic group) in our Reinforcement Learning experiments.

4.4 Sensitivity Analysis on Pearson Coefficient Threshold

In this section, we explore the influence of threshold value θ_{min}, as specified in Section 3.2.4, on the performance of multilinear regression models designed to predict future short-acting insulin doses and glucose levels. In other words, we perform feature selection by keeping the most relevant features based on Pearson Coefficient Coefficient. We have previously observed that performance improvements can be achieved using demographically clustered datasets. Therefore, our analysis in this Section focusses on training and evaluation datases from identical demographic groups.

4.4.1 Experimental Method for Identifying Best-Performing Features

In this section, we calculate the Pearson coefficient value for the best-performing features with respect to the target variable Short-acting insulin. By doing so, we establish the range of thresholds that will

be used in the Reinforcement Learning Section.

For this purpose, we conducted 10 independent experiments in which we examined the correlation between past independent variables of short-acting insulin doses and future dependent variables of short-acting insulin doses. In each test, we used data from 50 individuals who were randomly selected from the same demographic group. These groups are pre-processed as described in Section 3.2.5, and their features are subsequently scored following the methodology described in Section 3.2.4. Table 13 shows the best features that predict short-acting insulin doses for different groups: Non-Diabetic, Type I Diabetic and Type II Diabetic Patients. Each feature is evaluated on the basis of its Pearson correlation coefficient to determine its relevance in the accurate forecasting of future short-acting insulin needs. Note that Table 13 presents only one category of independent variables of previous short-acting insulin doses. It is observed that in the case of Past Short-Acting Insulin doses, small periods are more correlated with future short-acting insulin doses than larger periods (highlighted with bold text).

Dependent Variable Short-Acting Insulin Future Period (Minutes)	Independent Variable Short-Acting Insulin Past Period (Hours)	Pearson Correlation Coefficient
	Non Diabetic	
90	0.5	0.67
75	0.5	0.641
60	0.5	0.601
45	0.5	0.542
30	0.5	0.449
120	**0.5**	**0.724**
105	0.5	0.702
	Diabetic Type I	
90	1	0.801
75	1	0.81
60	0.8	0.817
45	0.8	0.823
30	0.2	0.832
15	**0.2**	**0.866**
120	1.2	0.784
105	1.2	0.792
	Diabetic Type II	
90	0.5	0.921
75	0.5	0.927
60	0.5	0.933
45	**0.5**	**0.937**
30	0.5	0.936
15	0.5	0.915
120	0.5	0.907
105	0.5	0.914

Table 13: Best Short-Acting Insulin Features for Future Short-Acting Insulin predictions using Pearson Correlation Coefficient

Since tables offer a limited view to observe the large number of correlations between predictor and target variables, we constructed a more compact way to observe multiple correlations. Figure 16

illustrates for the above experimental setting the correlations between future short-acting insulin target variables and predictive features of past Short-Acting Insulin.

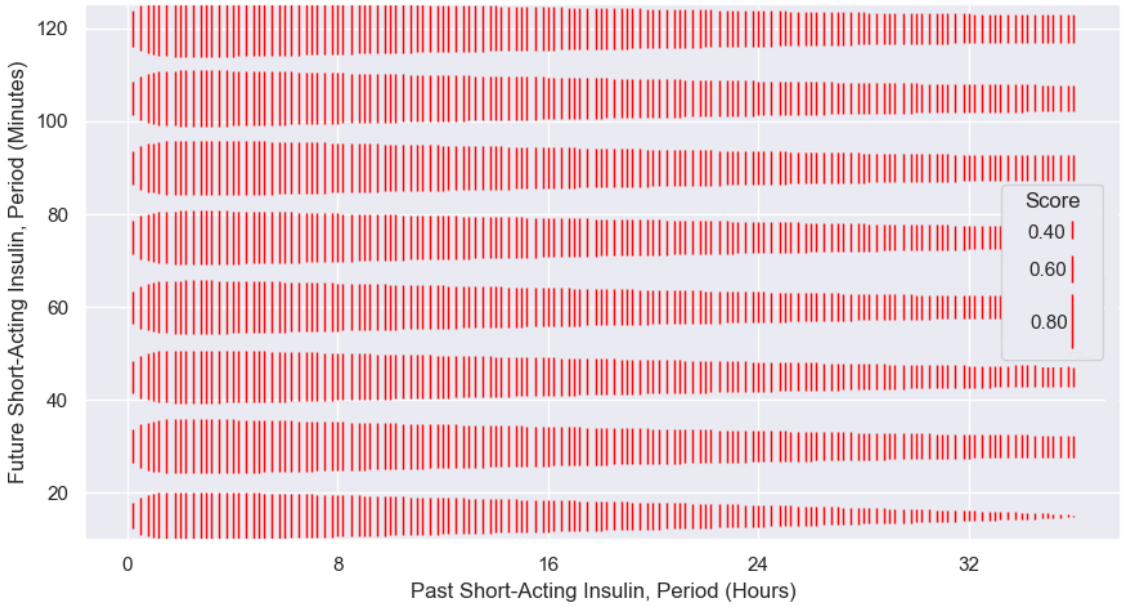

Figure 16: Correlations among all Future Short-Acting Insulin and Past Short-Acting Insulin Periods

It is observed again in more detail that small past periods demonstrate a stronger correlation than longer ones. Additionally, a consistent pattern emerges across all target variables when it comes to their correlation with predictive features.

In case of future glucose level predictions, the best performing predictor set is the one of past glucose levels. Figure 17 presents the correlations between future glucose level and past glucose level in all periods.

4.4.2 Effect of Threshold θ_{min} on Short-Acting Insulin Predictions

To explore how different values of threshold θ_{min} influence the predictive precision of a model, we conducted a sequence of discrete experiments. We separately examine the demographic groups Non Diabetic, Type I Diabetic, and Type II Diabetic patients.

In each experiment, we systematically varied θ_{min} at intervals of 0.05 from 0.1 to 0.95. The experimental settings are as follows: we averaged the evaluation results of 10 individual multilinear regression models per threshold value. For each model, we used a unique subset of 200 patients drawn from the examining demogrphic group. Furthermore, a separate set of 20 patients from the demogrphic group was randomly chosen to evaluate each model. It is worth noting that the patient sets used for training and evaluation of these models are different from those discussed in the feature

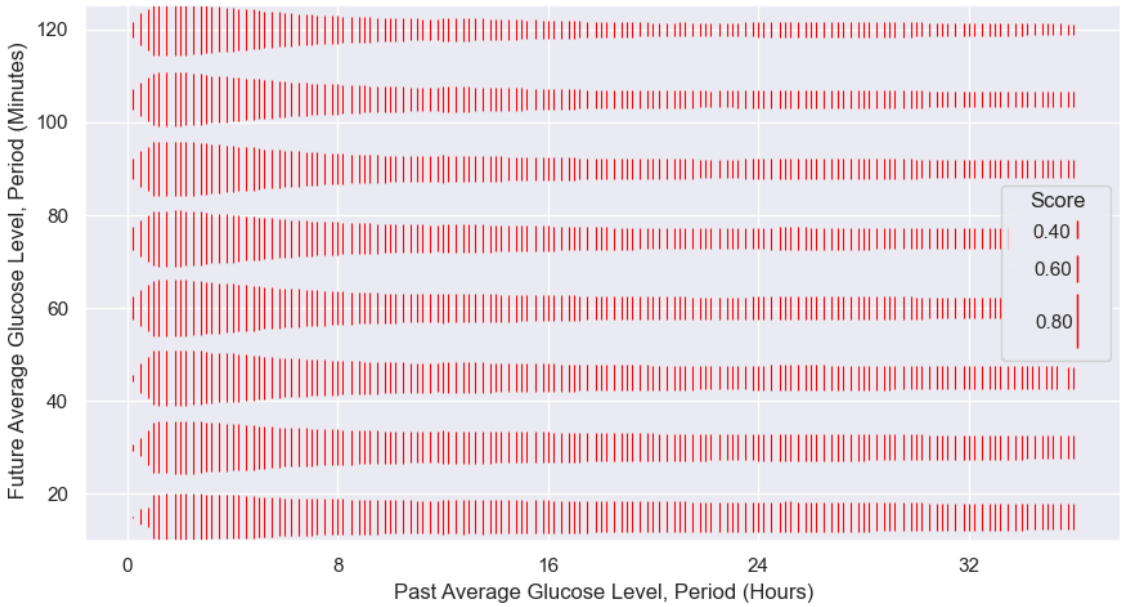

Figure 17: Correlations among all Future Glucose Level and Past Glucose Level Periods

pre-processing section, as detailed in Section 3.2.5.

Figures 24, 25 and 18 present the prediction performance for the demographic groups of Non-Diabetic, Type I Diabetic, and Type II Diabetic patients over Threshold θ_{min} values from 0.1 to 0.95 by intervals of 0.05. (see Appendix D for Figures 24, 25).

It is observed that in all cases, as the threshold θ_{min} increases, the Mean Absolute Error (MAE), Coefficient of Determination (R2), and the number of predictor features decrease. A decrease in MAE is expected, as more correlated information is used for training. The decrease in R2 values is also expected, since fewer data points are used for training.

4.5 Evaluation of Composite Multilinear Regression Models

The objective of this section is to assess the methodology introduced in Section 3.3. We aim to evaluate whether, using our previously developed experimental dataset and results, we can train an agent to construct a Composite Multilinear Regression Model to outperform a Single-Patient Regression Model built exclusively on the data of the individual patient under examination in terms of predictive accuracy. In addition, we will also discuss some practical choices we made to accelerate the training simulations.

In all of our experiments, we focus on predicting short-acting insulin doses for the next 1 hour. Future research could extend this focus to explore other target variables that we have defined.

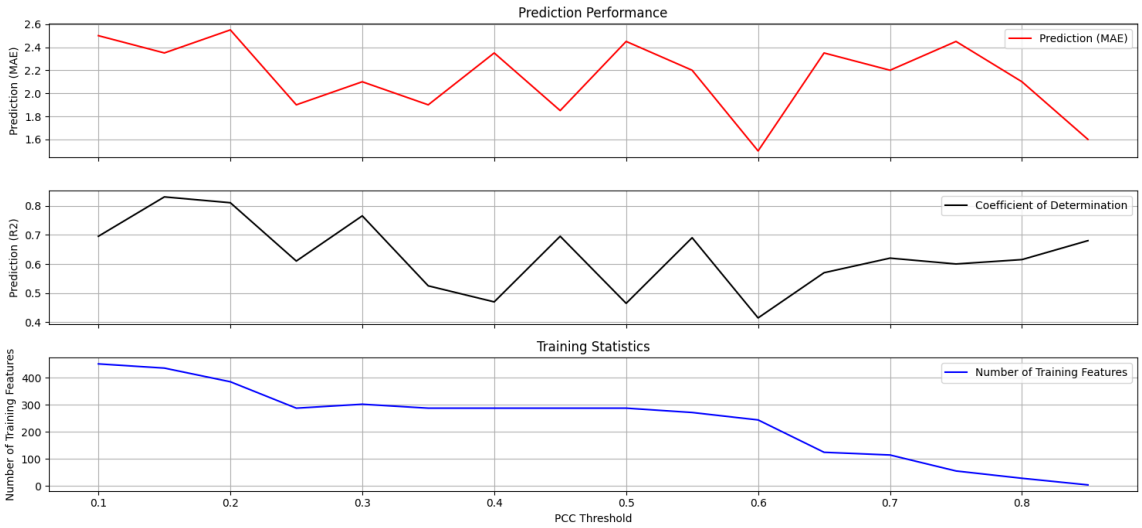

Figure 18: Performance of Multilinear Regression in 1 hour Short-Acting Insulin predictions over Threshold θ_{min} values from 0.1 to 0.95 by intervals of 0.05 in Diabetic TYPE II Patients

It is crucial to specify the portion of each patient's dataset that will be used to train the Single-Patient Models. In our experiments, we investigate the use of the first 33%, 66%, and 100% of the patient's data.

4.5.1 Generation of Single-Patient Multilinear Regression Models Collection

In order to accelerate the training speed of our Q-Network. A collection of pre-trained multilinear regression models was produced. The target variable in all models was selected to be the short-acting insulin doses for the upcoming hour.

Using the first 33%, 66% and 100% subset of the data from each 8,225 patients, we train and evaluate 9 different multilinear regression models, using PCC thresholds from (0.1) to (0.9) by intervals of 0.1. 80% of the data set subset was used for training and the rest 20% for evaluation. The patient data sets were filtered so that the glucose level 1 hour after insulin administration was in the target range of 80 to 180 mg/dL. Among the large number of generated models, we selectively exclude those with high-evaluation MAEs (> 15). Additionally, models are not produced when the PCC threshold for feature selection eliminates all features or when an insufficient number of training samples are available.

The final collection contains 42,458 models categorised by i) Patient's Subject ID, ii) Employed Percentage over the total of patient's dataset, and iii) PCC threshold value and iv) a flag to identify the best performing models with $MAE \geq 0$, $MAE < 5$, $R2 \geq 0.2$, and $R2 < 1$. Table 14 presents an overview of the model collection.

			MAE ≥ 0, $MAE < 5$, $R2 \geq 0.2$, and $R2 < 1$				
				MAE		R2	
ICU Time %	PCC Threshold	No of Models	No of Models	Mean Value	σ	Mean Value	σ
33%	0.1	1950	615	0.8	0.72	0.7	0.22
33%	0.2	1946	557	0.8	0.74	0.7	0.23
33%	0.3	1913	499	0.8	0.72	0.7	0.23
33%	0.4	1817	485	0.8	0.78	0.7	0.22
33%	0.5	1668	453	0.9	0.76	0.7	0.22
33%	0.6	1430	398	0.8	0.71	0.7	0.23
33%	0.7	1150	367	0.8	0.8	0.7	0.22
33%	0.8	795	265	0.8	0.75	0.7	0.21
33%	0.9	447	146	0.8	0.78	0.8	0.19
66%	0.1	2445	700	0.8	0.68	0.7	0.22
66%	0.2	2417	587	0.9	0.78	0.7	0.22
66%	0.3	2311	524	0.8	0.73	0.7	0.22
66%	0.4	2097	480	0.9	0.72	0.7	0.24
66%	0.5	1825	418	0.9	0.77	0.6	0.26
66%	0.6	1458	392	1.0	0.81	0.6	0.25
66%	0.7	1054	338	1.0	0.79	0.7	0.23
66%	0.8	678	235	0.9	0.72	0.7	0.22
66%	0.9	374	103	0.8	0.76	0.8	0.21
100%	0.1	2786	518	1.0	0.78	0.7	0.23
100%	0.2	2722	428	1.0	0.78	0.7	0.23
100%	0.3	2514	369	1.1	0.88	0.6	0.25
100%	0.4	2169	330	1.1	0.85	0.6	0.25
100%	0.5	1718	340	1.2	0.87	0.6	0.25
100%	0.6	1232	293	1.2	0.83	0.6	0.23
100%	0.7	801	211	1.0	0.68	0.6	0.22
100%	0.8	477	126	1.2	0.92	0.7	0.2
100%	0.9	264	55	0.8	0.66	0.8	0.19

Table 14: Statistics on Multilinear Regration Models for each patient seperatelly

Finally, we created a data structure in the form of a Python Pandas DataFrame, where each row includes patient demographic information, the four model's classifiers mentioned earlier, and the coefficients of the multilinear regression model. This data structure plays a crucial role in our RL training, as it allows us to easily identify patients similar to those being examined.

4.5.2 Training and Evaluation of Q-Networks

In this section, we conduct experiments to train and evaluate a set of agents designed to predict short-acting insulin doses for the next hour.

The agent's objective is to identify a set of patients similar to the one under examination and to determine a unique Pearson Correlation Coefficient (PCC) threshold value. This threshold will be applied to the training dataset of each individual patient within that similar set.

After calculating the population and threshold values, individual multilinear regression models are trained for each patient using the remaining features in each individual training. This collection is called the composite multilinear regression model or Multiple-Patients Regression Models. When it

comes time to determine the insulin dosage for a particular patient under examination, the predictive features for that specific moment are input into each model separately. Any features that are present in the patient being examined but missing from an individual model are excluded from that model's input. Conversely, if an individual model requires features that the patient under examination lacks, a value of zero is used to replace the missing feature. The final prediction is computed as the average value of all individual model predictions.

The baseline for performance comparison is a Single-Patient Regression Model, defined as a Multilinear Regression Model trained and evaluated using a subset of an individual patient's dataset. In our experimental approach, we utilize two distinct dataset sizes: one comprising first 33% of examining patient's data (with 26% for training and 7% for evaluation), and another using the full 100% (with 80% allocated for training and 20% for evaluation). Experimentation with the 66%, as referenced in Section 4.5.1, is reserved for future research.

During the training of the agent the coefficients of the Single-Patient Regression Model and a random population and threshold value are presented to the Q-Network. The output of the network represents an action on increasing or reducing the value of the population, and the threshold by one step. The final values for both the population and the threshold are determined when the agent chooses to keep these parameters constant across two consecutive steps.

The performance (MAE value) of the composite multilinear regression models is assessed after the agent's final action (the re-adjustment of population and threshold values) in the last step of each episode.

In each episode, a random patient is introduced by the Environment. For the purposes of our experimentation, we exclude patients for whom their Single-Patient Regression Model yielded MAE values greater than 5. This exclusion applies to both the 33% and 100% training sets. Specifically, out of 5,771 patients presented in the 33% training set, 4,811 were actually used for agent training. Similarly, for the 100% training set, 3,279 out of 4,257 patients were employed in the training process. Since, all relevant parameters are detailed in Section 3.3, we can proceed to present our findings.

Figure 19 provides an illustration of the training progress when using the training set 33%. In this figure, we gain insights into the magnitudes of cumulative rewards and the rate of epsilon decay. The equivalent figure for the 100% training set is located in the Appendix E.
The performance of the agent throughout the training process is depicted in Figure 27. We observe that the Composite Multilinear Regression Models consistently outperform the Single-Patient Model over the course of training. The corresponding figure for the 100% training set can be found in the Appendix E.

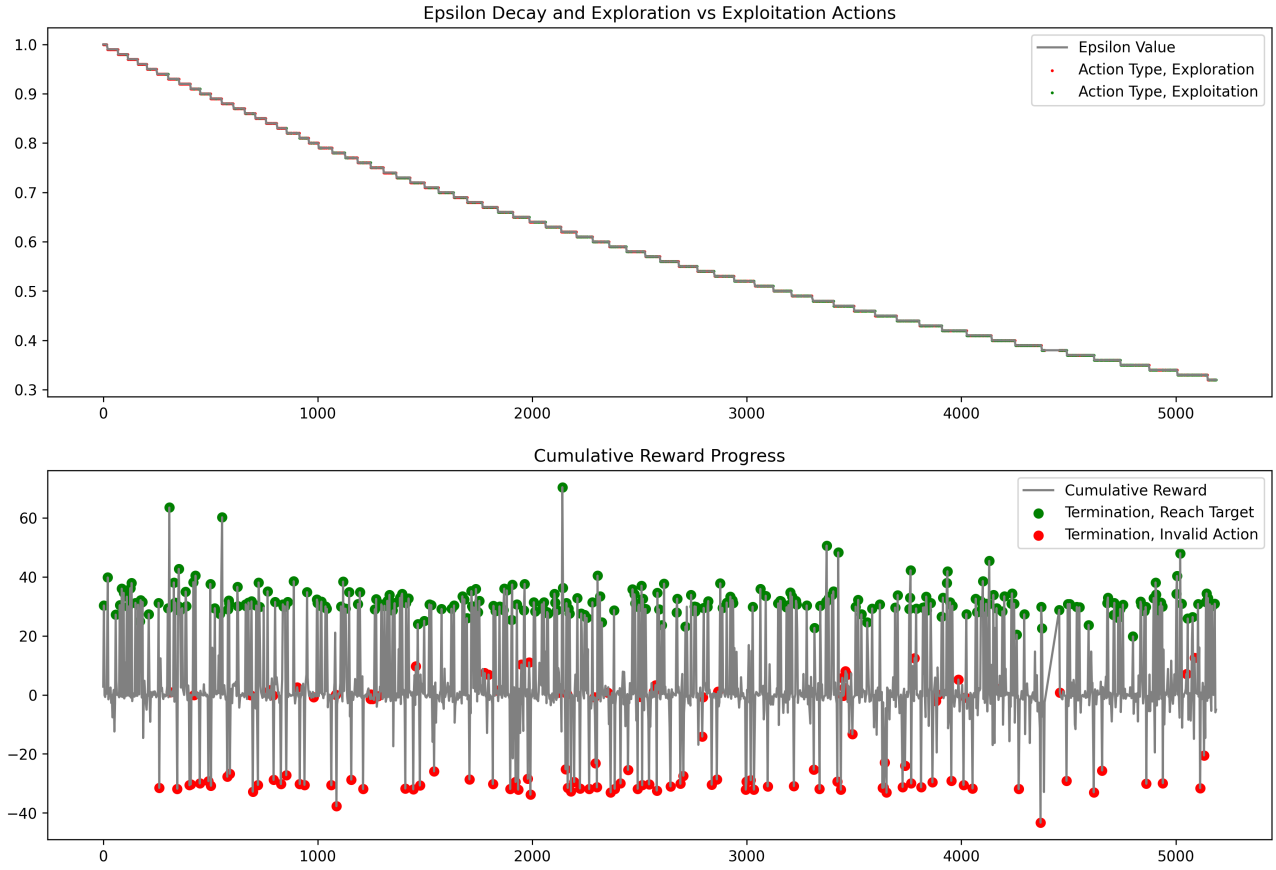

Figure 19: Epsilon Decay and Cumulative Reward during Training Progress using as Single-Patient Model Coefficients trained on first 33% of patient dataset

A more insightful way to assess whether an agent improves with an increasing number of episodes can be extracted from Figure 28. Each point in the graph represents the average value of the MAE measurements of the last N/x patients, where N is the total number of episodes and x the value of the x-axis. As can be observed, there is a notable improvement in performance, as evidenced by the reduction in MAE from 2.9 to 2.2. The corresponding figure for the 100% training set can be found in the Appendix E, in which we observe a similar improvement in performance.

Finally, we conduct a final evaluation to observe the performance of the agents in various diagnostic groups. Table 15 presents the average MAE values for both Single and Multiple-Patient Models, calculated over the last 1,000 patients presented during the training process. We find that in all cases, the Composite Models outperform the Single-Patient Models. Table 14 presents an overview of the model collection.

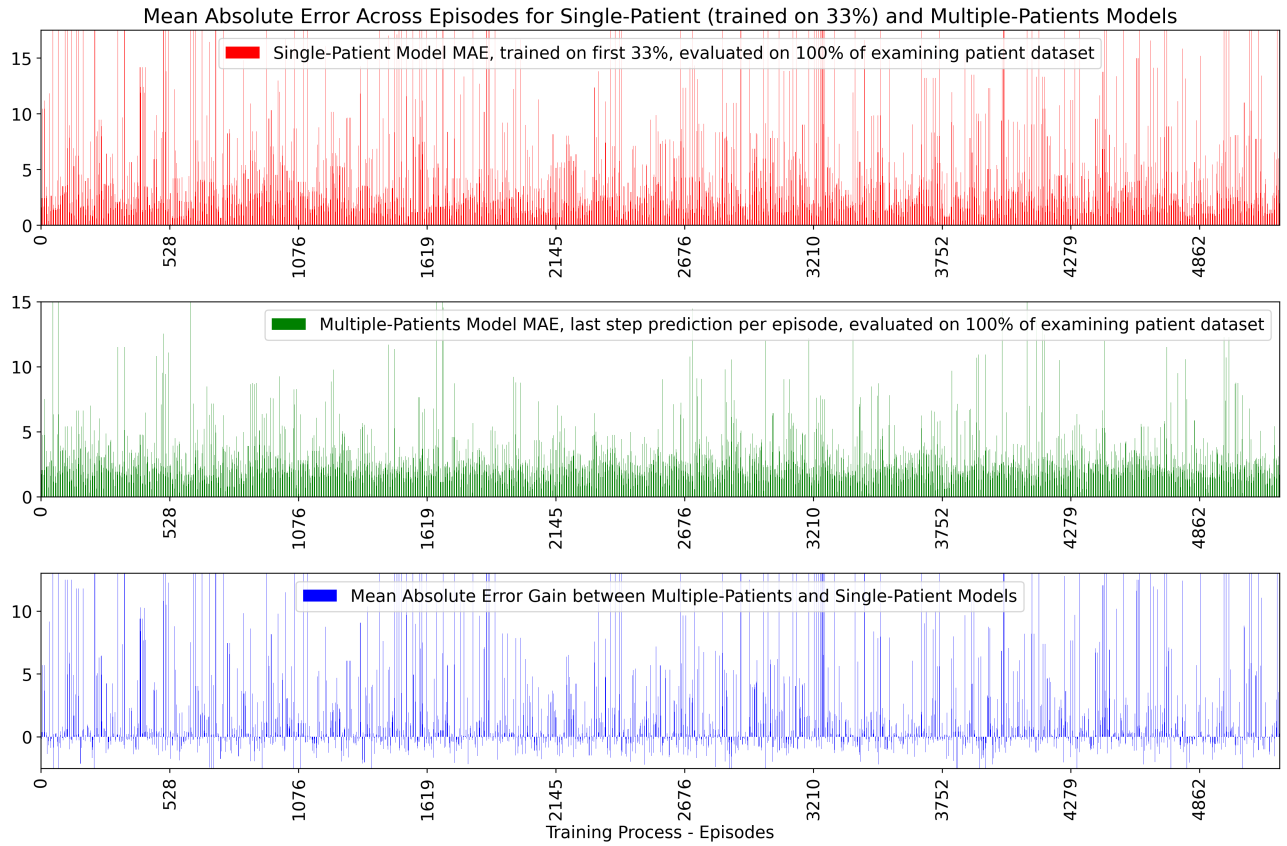

Figure 20: Mean Absolute Error Across Episodes for Single-Patient (trained on 33% of examining patient dataset) and Multiple-Patients Models

Diagnosis Group	Training Set %	Number of Examining Patients	Single-Patient Model (MAE)	Multiple-Patient Models (MAE)
Unable to Predict	33%	152	-	-
Non-Diabetic	33%	509	3.14	**2.33**
Diabetic Type I	33%	32	3.78	**3.36**
Diabetic Type II	33%	307	7.3	**3.68**
General Population	33%	848	4.67	**2.85**
Unable to Predict	100%	125	-	-
Non-Diabetic	100%	592	2.63	**1.99**
Diabetic Type I	100%	33	13.96	**5.17**
Diabetic Type II	100%	250	8.39	**3.61**
General Population	100%	875	4.7	**2.57**

Table 15: Evaluation on the last 1000 Patients presented to the Agent

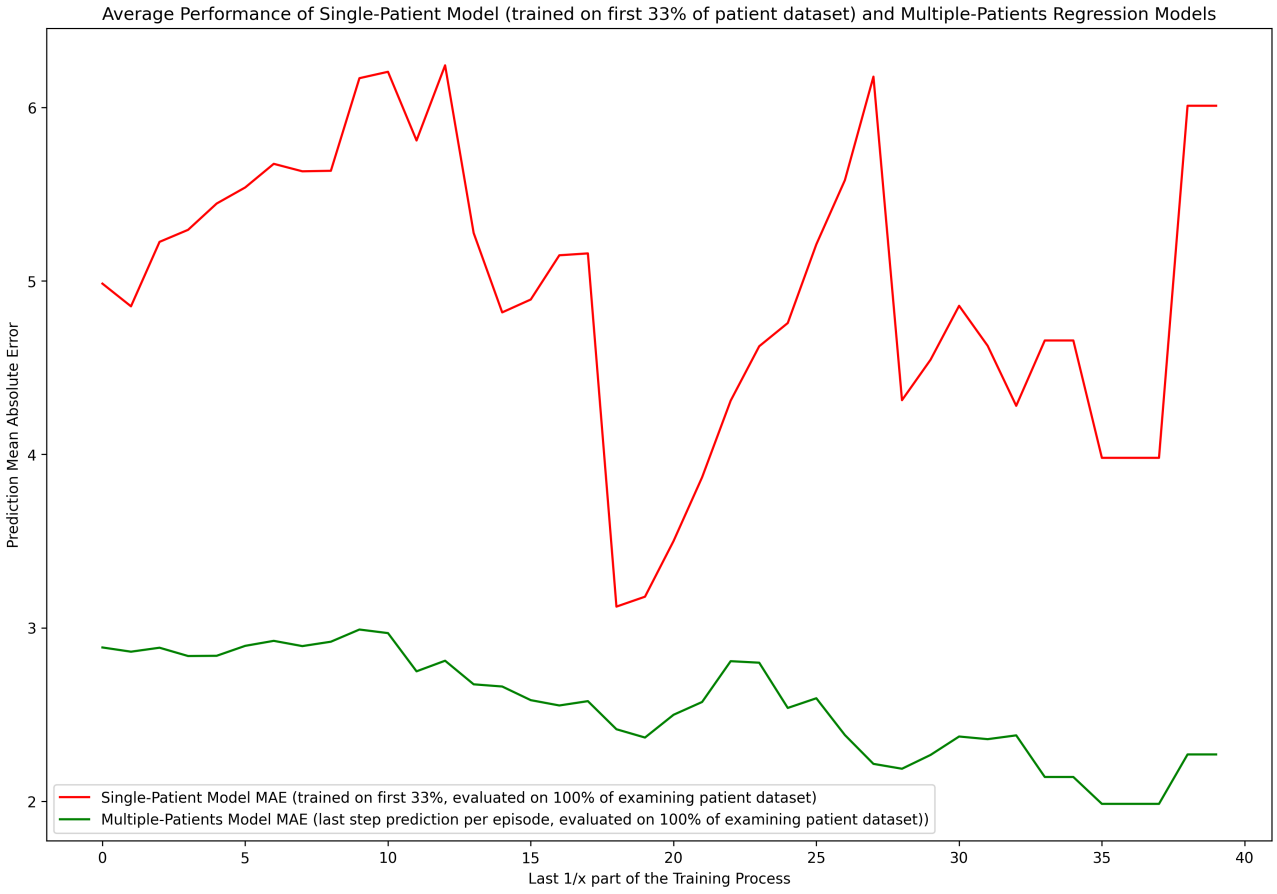

Figure 21: Average Performance of Single-Patient Model (Trained on Initial 33% of Patient Data) Versus Multiple-Patients Regression Models Across 4,811 Patients Spanning All Demographic Groups

5 Conclusion and Future Work

In this chapter, we aim to include the derived conclusions and briefly discuss potential aspects of our research work. At the last section we descuss how the currently achieved results may be improved in future academic endeavours.

5.1 Conclusions

In our work, we successfully propose and evaluate a methodology for predicting future insulin levels for both diabetic and non-diabetic patients in the ICU. Utilizing Multilinear Regression models as the foundational building blocks and Deep Q-learning as a tool, we are able to produce agents capable of constructing Composite Multilinear Regression Models that outperform Single-Patient Regression Models.

To achieve this, we evaluated the extent to which the defined features could predict future insulin doses in various feature and instance filtering scenarios, using the specified evaluation metrics. For these experiments, we constructed a dataset based on MIMIC III [23], which contains a predefined set of predictive and target variables across a large number of insulin administrations for a demographically diverse set of patients.

The preliminary experiments included Learning Curve Analysis on randomly selected and demographically identical patients, as well as feature selections using Pearson Correlation Coefficients. These experiments indicated that training datasets with demographically uniform patients produce more accurate predictions. Additionally, we drew some conclusions regarding the best-performing features for specific targets. For instance, short-acting insulin targets and predictive features of past short-acting insulin doses are highly correlated.

Finally, we trained Deep Q-Networks under various training scenarios and concluded that an agent can be trained to produce Composite Multilinear Regression Models with superior predictive accuracy compared to Single-Patient Regression Models built exclusively on the data of the patient under examination. For instance, in the case of Non-Diabetic patients, Composite Multilinear Regression Models yielded an average MAE value of 2.33, as opposed to 3.14 for Single-Patient Models. Similarly, in the Type II Diabetic group, the measurements were 3.68 compared to 7.3, and for the General Population, they were 2.85 compared to 4.67.

5.2 Future Work

There several ways that the work of this thesis can be expanded. In this section we summarize some ideas. In regard to the experimentation dataset, different sets of Dependent and Independent Variables could be examined under a different time step than the selected 15 minutes. Additionally, parameters already present in the current version could be investigated further. For instance, we might explore whether patients who passed away in the ICU exhibit distinct patterns in insulin dosing and glucose measurements compared to the general population. We could also examine if the length of stay introduces bias, particularly due to potential repetitive patterns in treatments during extended stays. Finally, the subcategories listed in Table 16 could be used to filter specific subgroups.

Another point of interest is the selection of the insulin administration strategy. In our current work, multilinear regression models are trained to predict insulin dosages that aim to maintain glucose levels within the target range of 80 to 180 mg/dL. Considering that ICU patients typically have regular feeding schedules, we may explore filtering training instances based on the time of day in future work emulating Basal Bolus Insulin Therapy.

In regard to the Learning Curve Analysis, our current work is limited to 1 and 2 hour Short-Acting Insulin predictions and 1 and 2 hours future glucose levels. In future work, we can expanded our investigate to include more combinations of future insulin time and types which are already inludes in ohr experimentation dataset.

Also In the evaluation metrics, it is important to include the total number of times the glucose level is within the normal range. Finally, our Reinforcement Learning work serves as an excellent starting point for investigating various neural network architectures and simulation hyperparameters. Additionally, longer simulation times should be executed to draw conclusions about the final convergence point, as current simulations are abbreviated due to time constraints.

References

[1] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[2] European Medicines Agency. Humalog, inn-insulin lispro, 2023. URL: https://www.ema.europa.eu/en/documents/product-information/humalog-epar-product-information_en.pdf.

[3] European Medicines Agency. Humulin r, datasheet, 2023. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2013/018781s121lbl.pdf.

[4] European Medicines Agency. Lantus glargine, datasheet, 2023. URL: https://www.ema.europa.eu/en/documents/product-information/lantus-epar-product-information_en.pdf.

[5] Ethem Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2020.

[6] Simon R. Heller and. A summary of the ADVANCE trial. *Diabetes Care*, 32(suppl_2):S357–S361, nov 2009. doi:10.2337/dc09-s339.

[7] Novo Nordisk AS. Novolin r, datasheet, 2023. URL: https://www.novo-pi.com/novolinr.pdf.

[8] AuthorFirstName AuthorLastName. Deep reinforcement learning: Emerging trends in macroeconomics and future prospects, 2023. URL: https://www.imf.org/en/Publications/WP/Issues/2022/12/16/Deep-Reinforcement-Learning-Emerging-Trends-in-Macroeconomics-and-Future-Prospects-

[9] Marie Aare Bentsen, Zaman Mirzadeh, and Michael W. Schwartz. Revisiting how the brain senses glucose—and why. *Cell Metabolism*, 29(1):11–17, jan 2019. doi:10.1016/j.cmet.2018.11.001.

[10] Cari Berget, Laurel H. Messer, and Gregory P. Forlenza. A clinical overview of insulin pump therapy for the management of diabetes: Past, present, and future of intensive therapy. *Diabetes Spectrum*, 32(3):194–204, aug 2019. doi:10.2337/ds18-0091.

[11] R N Bergman, Y Z Ider, C R Bowden, and C Cobelli. Quantitative estimation of insulin sensitivity. *American Journal of Physiology-Endocrinology and Metabolism*, 236(6):E667, jun 1979. doi:10.1152/ajpendo.1979.236.6.e667.

[12] Wendy C. Bevier, Howard C. Zisser, Lois Jovanovič, Daniel A. Finan, Cesar C. Palerm, Dale E. Seborg, and Francis J. Doyle. Use of continuous glucose monitoring to estimate

insulin requirements in patients with type 1 diabetes mellitus during a short course of prednisone. *Journal of Diabetes Science and Technology*, 2(4):578–583, jul 2008. doi:10.1177/193229680800200408.

[13] Diabetes Teaching Center. Calculating insulin dose, March 2023. URL: https://diabetesteachingcenter.ucsf.edu/calculating-insulin-dose.

[14] N. R. Draper and H. Smith. *Applied Regression Analysis*. John Wiley & Sons, 1998.

[15] Marius A. Eriksen. Model predictive control for closed-loop insulin delivery. https://gist.github.com/mariusae, 2018.

[16] R. L. Fleurence, L. H. Curtis, R. M. Califf, R. Platt, J. V. Selby, and J. S. Brown. Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association*, 21(4):578–582, jul 2014. doi:10.1136/amiajnl-2014-002747.

[17] U.S. Food and Drug Administration. Humulin n nph, datasheet, 2023. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2013/018781s121lbl.pdf.

[18] Peter Foy. Deep reinforcement learning: Guide to deep q-learning, 2023. URL: https://www.mlq.ai/deep-reinforcement-learning-q-learning/.

[19] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[20] Today's Hospitalist. Tips for calculating a total daily dose of insulin, August 2007. URL: https://www.todayshospitalist.com/tips-for-calculating-a-total-daily-dose-of-insulin/.

[21] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. Springer New York, 2013. . doi:10.1007/978-1-4614-7138-7.

[22] Alistair Johnson, Tom Pollard, and Roger Mark. Mimic-iii clinical database, 2020. doi:10.13026/C2XW26.

[23] Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1), may 2016. doi:10.1038/sdata.2016.35.

[24] Vijay Kanade. What is the markov decision process? definition, working, and examples, 2022. URL: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-markov-decision-process/.

[25] Saira Khan, Sherita Hill Golden, and Nestoras Mathioudakis. Associations between home insulin dose adjustments and glycemic outcomes at hospital admission. *Diabetes Research and Clinical Practice*, 127:51–58, may 2017. doi:10.1016/j.diabres.2017.02.003.

[26] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[27] Narendra Kotwal and Aditi Pandit. Variability of capillary blood glucose monitoring measured on home glucose monitoring devices. *Indian Journal of Endocrinology and Metabolism*, 16(8):248, 2012. doi:10.4103/2230-8210.104052.

[28] K.Suresh, O.Obulesu, and B. Venkata Ramudu. Diabetes prediction using machine learning techniques. *Helix*, 10(02):136–142, Apr 2020.

[29] Pawel Ladosz, Lilian Weng, Minwoo Kim, and Hyondong Oh. Exploration in deep reinforcement learning: A survey. 2022. doi:10.48550/ARXIV.2205.00824.

[30] E.D. Lehmann and T. Deutsch. A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. *Journal of Biomedical Engineering*, 14(3):235–242, may 1992. doi:10.1016/0141-5425(92)90058-s.

[31] Xiran Liu, Ivana Jankovic, and Jonathan H Chen. Predicting inpatient glucose levels and insulin dosing by machine learning on electronic health records. mar 2020. doi:10.1101/2020.03.02.20029017.

[32] Chiara Dalla Man, Francesco Micheletto, Dayu Lv, Marc Breton, Boris Kovatchev, and Claudio Cobelli. The UVA/PADOVA type 1 diabetes simulator. *Journal of Diabetes Science and Technology*, 8(1):26–34, jan 2014. doi:10.1177/1932296813514502.

[33] mayoclinic.org. What are normal blood sugar levels? – diabetic me., March 2023. URL: https://diabeticme.org/learning-center/health-guides/what-are-normal-blood-sugar-levels/.

[34] Minh Nguyen, Ivana Jankovic, Laurynas Kalesinskas, Michael Baiocchi, and Jonathan H Chen. Machine learning for initial insulin estimation in hospitalized patients. *Journal of the American Medical Informatics Association*, 28(10):2212–2219, jul 2021. doi:10.1093/jamia/ocab099.

[35] National Library of Medicine. Type 1 diabetes: Types of insulin therapy, June 2017. URL: https://www.ncbi.nlm.nih.gov/books/NBK279338/.

[36] National Library of Medicine. Icd-9-cm diagnosis codes defining diabetes, May 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK368403/table/sb203.t5/.

[37] Pasquale Palumbo, Susanne Ditlevsen, Alessandro Bertuzzi, and Andrea De Gaetano. Mathematical modeling of the glucose–insulin system: A review. *Mathematical Biosciences*, 244(2):69–81, aug 2013. doi:10.1016/j.mbs.2013.05.006.

[38] Francisco J Pasquel, M Cecilia Lansang, Ketan Dhatariya, and Guillermo E Umpierrez. Management of diabetes and hyperglycaemia in the hospital. *The Lancet Diabetes &amp Endocrinology*, 9(3):174–188, mar 2021. doi:10.1016/s2213-8587(20)30381-8.

[39] PCORnet. Patient-centered clinical research network, 2020. Accessed: 15.08.2023. URL: https://pcornet.org/data/.

[40] Ishwarlal Jialal Rajeev Goyal. Type 2 diabetes, May 2023. URL: https://pubmed.ncbi.nlm.nih.gov/30020625/.

[41] Joseph Lee Rodgers and W. Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, feb 1988. doi:10.1080/00031305.1988.10475524.

[42] RimaB Shah, ViralN Shah, Manhar Patel, and DavidM Maahs. Insulin delivery methods: Past, present and future. *International Journal of Pharmaceutical Investigation*, 6(1):1, 2016. doi:10.4103/2230-973x.176456.

[43] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction*. MIT press, 2018.

[44] Miguel Tejedor, Ashenafi Zebene Woldaregay, and Fred Godtliebsen. Reinforcement learning application in diabetes blood glucose control: A systematic review. *Artificial Intelligence in Medicine*, 104:101836, apr 2020. doi:10.1016/j.artmed.2020.101836.

[45] Medical News Today. An overview of insulin, May 2023. URL: https://www.medicalnewstoday.com/articles/323760#types-of-insulin.

[46] Guillermo E. Umpierrez, Dawn Smiley, Ariel Zisman, Luz M. Prieto, Andres Palacio, Miguel Ceron, Alvaro Puig, and Roberto Mejia. Randomized study of basal-bolus insulin therapy in the inpatient management of patients with type 2 diabetes (RABBIT 2 trial). *Diabetes Care*, 30(9):2181–2186, sep 2007. doi:10.2337/dc07-0295.

[47] Andre Violante. Simple reinforcement learning: Q-learning. Towards Data Science, Mar 2019. URL: https://towardsdatascience.com/simple-reinforcement-learning-q-learning-fcddc4b6fe56.

Appendices: Assorted Tables and Figures

A Population Distribution Across Diabetes Subcategories

Category	Subategory	Population
TYPE II	Diabetes Without Mention Of Complication	3091
TYPE II	Diabetes With Neurological Manifestations	418
TYPE II	Diabetes With Renal Manifestations	262
TYPE II	Diabetes With Other Specified Manifestations	171
TYPE II	Diabetes With Ophthalmic Manifestations	158
TYPE II	Diabetes With Ketoacidosis	103
TYPE II	Diabetes With Neurological Manifestations	98
TYPE II	Diabetes With Renal Manifestations	57
TYPE II	Diabetes With Other Specified Manifestations	56
TYPE II	Diabetes With Peripheral Circulatory Disorders	41
TYPE II	Diabetes With Ophthalmic Manifestations	40
TYPE II	Diabetes With Unspecified Complication	24
TYPE II	Diabetes With Hyperosmolarity	24
TYPE II	Diabetes With Peripheral Circulatory Disorders	15
TYPE II	Diabetes With Unspecified Complication	11
TYPE II	Diabetes With Hyperosmolarity	9
TYPE II	Diabetes With Other Coma	3
TYPE II	Diabetes With Other Coma	1
TYPE I	Diabetes With Ketoacidosis	176
TYPE I	Diabetes With Neurological Manifestations	77
TYPE I	Diabetes With Ophthalmic Manifestations	54
TYPE I	Diabetes With Neurological Manifestations	49
TYPE I	Diabetes Without Mention Of Complication	48
TYPE I	Diabetes With Renal Manifestations	45
TYPE I	Diabetes With Renal Manifestations	37
TYPE I	Diabetes With Ophthalmic Manifestations	32
TYPE I	Diabetes With Other Specified Manifestations	22
TYPE I	Diabetes With Other Specified Manifestations	20
TYPE I	Diabetes Without Mention Of Complication	11
TYPE I	Diabetes With Peripheral Circulatory Disorders	7
TYPE I	Diabetes With Peripheral Circulatory Disorders	4
TYPE I	Diabetes With Unspecified Complication	3
TYPE I	Diabetes With Other Coma	3
TYPE I	Diabetes With Unspecified Complication	1
TYPE I	Diabetes With Ketoacidosis	1
TYPE I	Diabetes With Hyperosmolarity	1

Table 16: Population Distribution Across Diabetes Subategories of the 8,225 patients dataset, based on ICD9 Prefix 250

B Independent Variables Descriptive Statistics

Feature Description	Mean	σ	IQR
Average Glucose Level last 3 Hours (mg/dL)	157.67	59.68	55.5
Average Glucose Level last 6 Hours (mg/dL)	159.62	60.44	56.8
Average Glucose Level last 12 Hours (mg/dL)	161.22	57.97	55.6
Average Glucose Level last 24 Hours (mg/dL)	162.91	54.47	54.5
Standard Deviation of Glucose Level last 3 Hours (mg/dL)	19.9	16.41	14.2
Standard Deviation of Glucose Level last 6 Hours (mg/dL)	26.37	22.11	19.7
Standard Deviation of Glucose Level last 12 Hours (mg/dL)	32.73	26.25	24.4
Standard Deviation of Glucose Level last 24 Hours (mg/dL)	38.25	28.15	29.4
Total Insulin last 3 Hours (Units)	18.62	26.03	15.02
Total Insulin last 6 Hours (Units)	31.4	43.85	28.62
Total Insulin last 12 Hours (Units)	53.61	75.29	52.06
Total Insulin last 24 Hours (Units)	88.77	132.26	93.52
Rapid Acting Insulin last 3 Hours (Units)	9.13	8.41	8.0
Rapid Acting Insulin last 6 Hours (Units)	8.76	8.38	9.0
Rapid Acting Insulin last 12 Hours (Units)	11.54	11.1	11.0
Rapid Acting Insulin last 24 Hours (Units)	16.35	16.02	14.0
Short Acting Insulin last 3 Hours (Units)	16.79	22.79	13.37
Short Acting Insulin last 6 Hours (Units)	29.4	40.18	25.87
Short Acting Insulin last 12 Hours (Units)	51.08	70.01	48.0
Short Acting Insulin last 24 Hours (Units)	85.06	121.12	88.02
Long Acting Insulin last 3 Hours (Units)	30.23	23.99	28.0
Long Acting Insulin last 6 Hours (Units)	30.48	24.82	30.0
Long Acting Insulin last 12 Hours (Units)	32.49	26.96	28.0
Long Acting Insulin last 24 Hours (Units)	38.12	41.91	26.0

Table 17: *Mean Value*, *Standard Deviation* and *Interquartile Range* for Independent Variables last 3, 6, 12 and 24 hours

C Toy Example Graphs

C.1 Example Patient's Next-1-Hour Short-Acting Insulin Administistrations

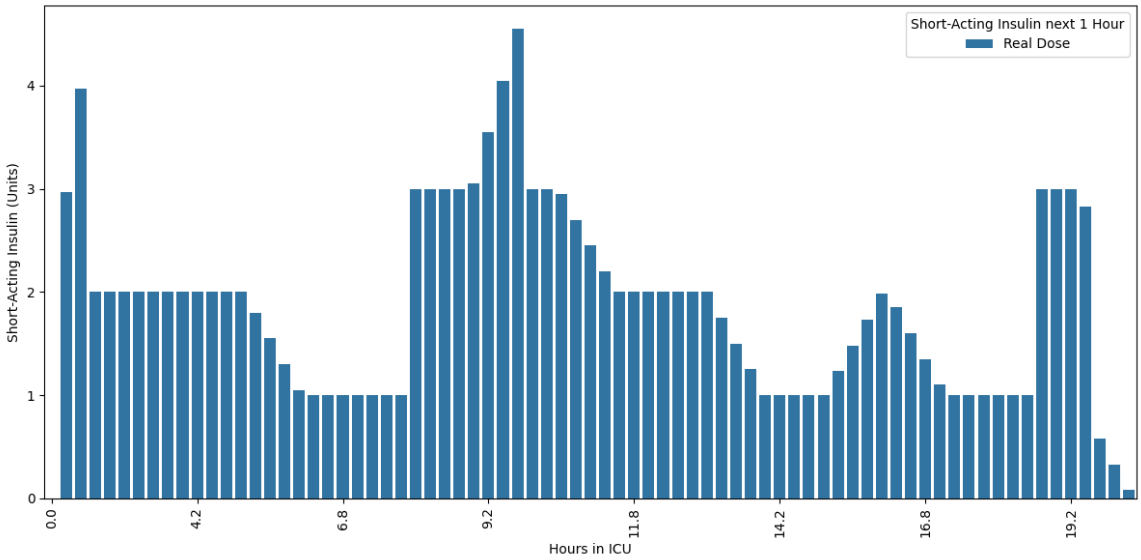

Figure 22: Example Patient's Next-1-Hour Short-Acting Insulin Administistrations Doses over Time

C.2 Single Patient Multilinear Regression Model Evaluation

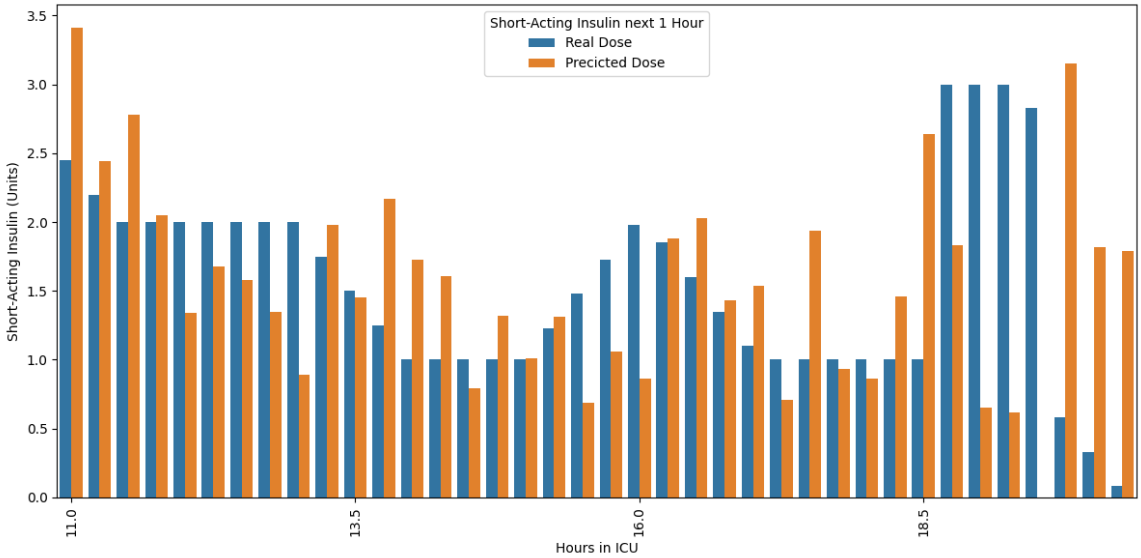

Figure 23: Real and Predicted Next-1-Hour Short-Acting Insulin Doses using Multilinear Regression Model trained on 50% of patient's data over Time

D Sensitivity Analysis on Pearson Coefficient Threshold

D.1 Sensitivity Analysis of θ_{min} on Non-Diabetic Patients

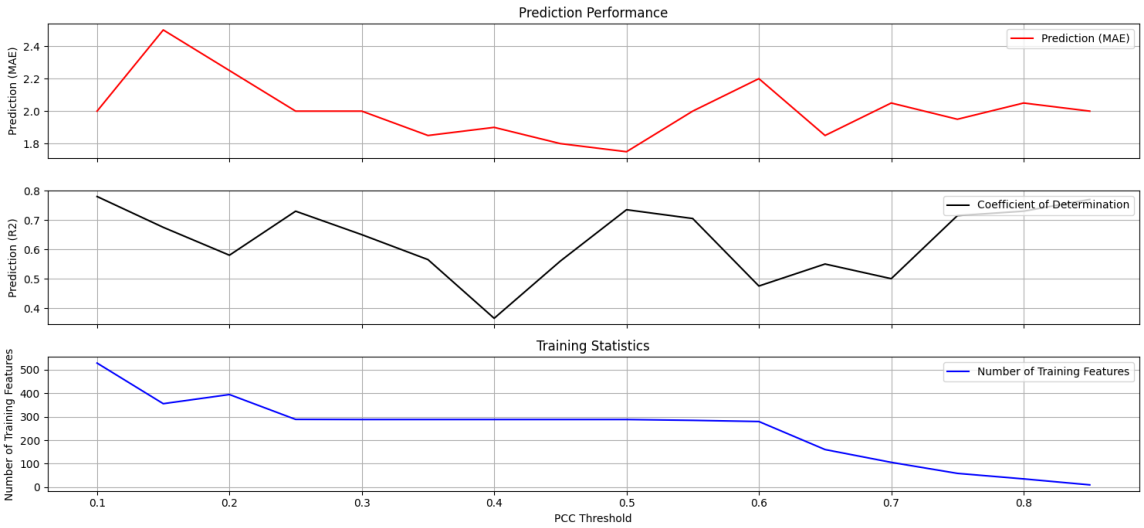

Figure 24: Performance of Multilinear Regression in 1 hour Short-Acting Insulin predictions over Threshold θ_{min} values from 0.1 to 0.95 by intervals of 0.05 in Non-Diabetic Patients

D.2 Sensitivity Analysis of θ_{min} on Diabetic TYPE I Patients

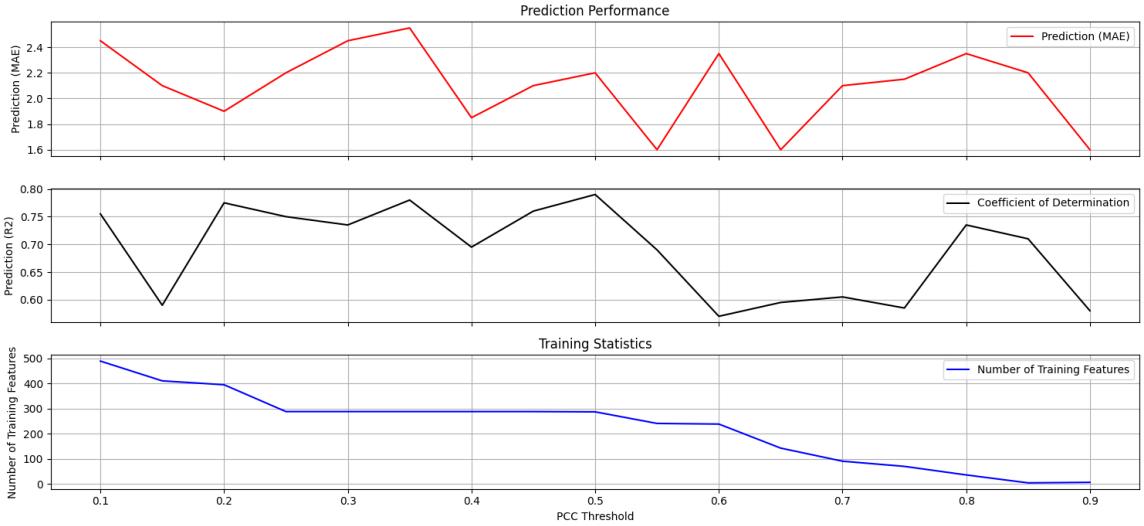

Figure 25: Performance of Multilinear Regression in 1 hour Short-Acting Insulin predictions over Threshold θ_{min} values from 0.1 to 0.95 by intervals of 0.05 in Diabetic TYPE I Patients

E Training and Evaluation of Q-Networks

E.1 Q-Network Training Progress Using Single-Patient Models employing 100% of Examining Patient Dataset

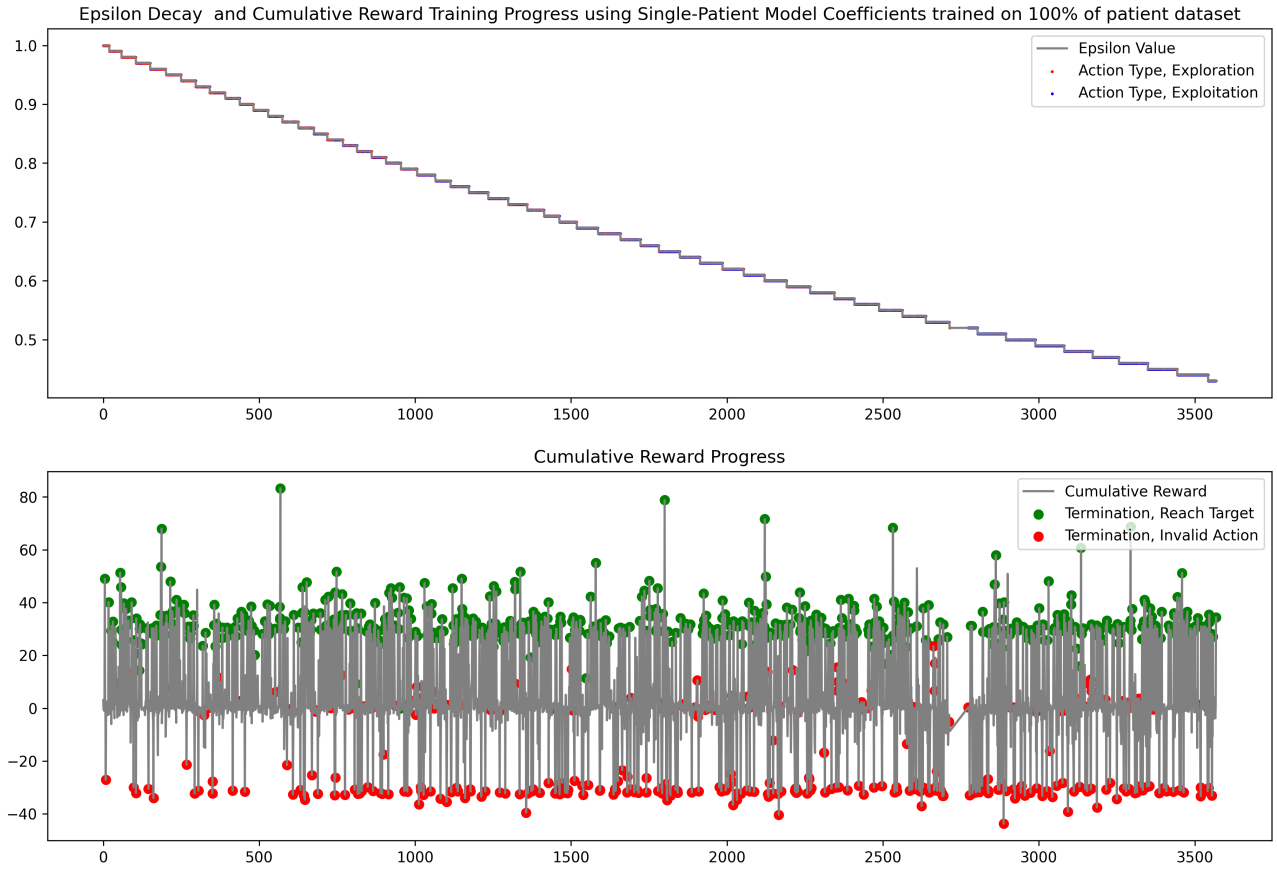

Figure 26: Epsilon Decay and Cumulative Reward Training Progress using Single-Patient Model Coefficients trained on 80% of patient dataset

E.2 Q-Network Evaluation Across Episodes Using Single-Patient Models employing 100% of Examining Patient Dataset

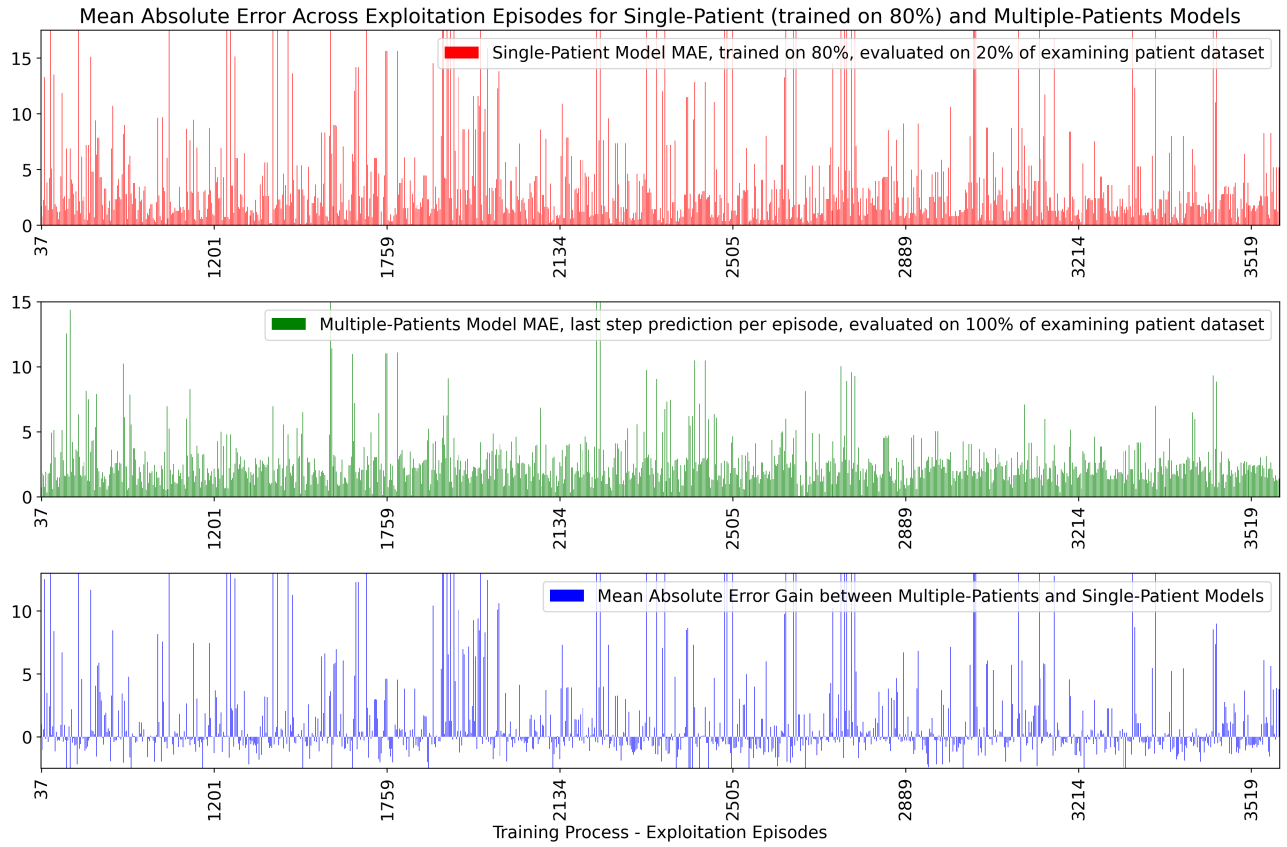

Figure 27: Mean Absolute Error Across Episodes for Single-Patient (trained on 80% of examining patient dataset) and Multiple-Patients Models

E.3 Average Performance of Single-Patient Models employing 100% of Examining Patient Dataset Versus Multiple-Patients Regression Models

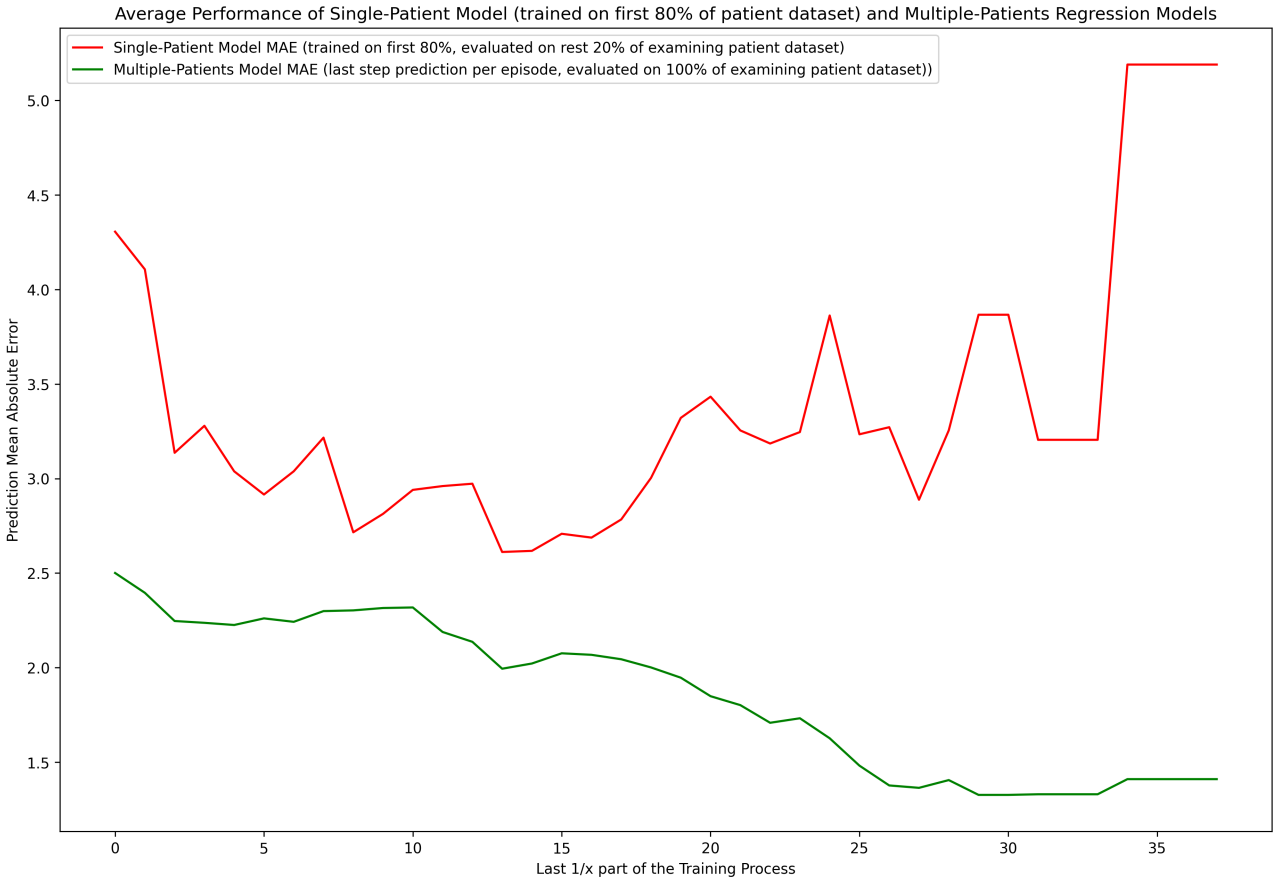

Figure 28: Average Performance of Single-Patient Model (Trained on 80% of Patient Data) Versus Multiple-Patients Regression Models Across 3,279 Patients Spanning All Demographic Groups