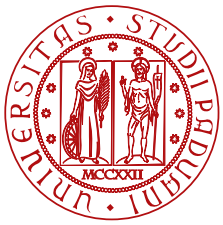


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DEPARTMENT OF INFORMATION ENGINEERING

MASTER'S DEGREE COURSE IN BIOENGINEERING

Time Series-to-Image Encoding approaches for Blood Glucose Levels Forecasting in Type 1 Diabetes using Deep Learning Models

Supervisor
Prof. Andrea Facchinetti

Candidate
Jessica Bruttomesso

Co-Supervisor
Ph.D. Francesco Prendin

ACADEMIC YEAR 2024-2025

24/02/2025

Abstract

Type 1 diabetes (T1D) is a chronic autoimmune disease characterized by the destruction of pancreatic β-cells, leading to impaired insulin production and to elevated blood glucose (BG) levels. Without an accurate glycemic control, individuals with T1D can face frequent episodes of hyperglycemia (BG > 180 mg/dL) and hypoglycemia (BG < 70 mg/dL), which can result in severe complications such as cardiovascular disease, kidney failure, neuropathy, and blindness.

An accurate management of T1D requires a complex and burdensome daily routine: frequent BG monitoring, carbohydrate counting, precise insulin administration, physical exercise and strict dietary management. Continuous Glucose Monitoring (CGM) are minimally invasive sensors that provide real-time glucose readings, typically every five minutes, as well as visual and acoustic alarms to warn patients of hypo- or hyperglycemic episodes. However, knowing ahead-in-time if glucose is approaching harmful levels would be much more helpful as it would enable proactive interventions.

BG forecasting is a well-established research area in the field of T1D with a vast body of literature exploring different methodologies to improve prediction accuracy. A wide range of models, from traditional statistical approaches to advanced machine learning and deep learning techniques. Despite these advancements, the development of BG forecasting algorithms is still an open problem, and its clinical applicability remains an ongoing challenge due to factors such as variability in individual responses, inaccuracies in sensor measurements, and the difficulty in capturing the full complexity of glucose dynamics.

One emerging trend in time series forecasting involves transforming sequential data into images, enabling the advanced Convolutional Neural Networks (CNN) which proves promising performance in the field of computer vision techniques for improved prediction accuracy. This novel approach has already demonstrated superior performance in various fields such as financial forecasting, weather pattern classification, and medical diagnostics.

This thesis explores the application of time series-to-image encoding for BG prediction in T1D. Particularly, glucose time series data are converted into images through three encoding techniques: Recurrence Plot (RP), Gramian Angular Field (GAF), and Markov Transition Field (MTF). These representations are then fed into a Convolutional Neural Network (CNN) trained to predict BG levels across various prediction horizons. To evaluate this approach, we conduct a comparative analysis of different input lengths and prediction horizons, benchmarking a Long Short-Term Memory (LSTM) network, which directly processes glucose data. The dataset used in this research consists of CGM data from 100 virtual individuals monitored for 30 days, generated using the UVA/Padova Type 1 Diabetes Mellitus Simulator (T1DMS).

The thesis is organized as follows: a brief introduction on T1D complications, management, technologies and an overview of the main state-of-art predictive algorithms is illustrated in Chapter 1.

Chapter 2 presents the analysis of the dataset used in this work for the assessment of the forecasting methodologies.

Chapter 3 provides a detailed description of three time-series-to-image encoding techniques explored in this study: Recurrence Plot (RP), Gramian Angular Field (GAF) and Markov Transition Field (MTF). Moreover it presents the predictive algorithms developed for BG forecasting: an Autoregressive model (AR), a Long Short-Term Memory (LSTM) network, and a Convolutional Neural Network (CNN).

Chapter 4 details the results, providing important insights into the predictive capabilities of CNN models trained on images, as well as a comparison with the predictive performance with the benchmark LSTM model trained on time series data.

Chapter 5 analyzes the main findings obtained from the study, exploring their implications and comparing the performance of the models developed

Finally, Chapter 6 summarizes the conclusions of the study, highlighting some limitations and outlining promising directions for future research.

Sommario

Il diabete di tipo 1 (T1D) è una malattia autoimmune cronica caratterizzata dalla distruzione delle cellule β del pancreas, che porta a una produzione insufficiente di insulina e a livelli elevati di glucosio nel sangue. Senza un controllo glicemico adeguato, le persone con T1D possono affrontare frequenti episodi di iperglicemia (BG > 180 mg/dL) e ipoglicemia (BG < 70 mg/dL), che possono portare a gravi complicazioni come malattie cardiovascolari, insufficienza renale, neuropatia e cecità.

Una gestione adeguata del T1D richiede una routine quotidiana complessa e impegnativa: monitoraggio frequente della glicemia, conteggio dei carboidrati, somministrazione precisa dell'insulina, esercizio fisico e una dieta rigorosa. I sensori di monitoraggio continuo della glicemia (CGM) sono sensori minimamente invasivi che forniscono letture in tempo reale della glicemia, tipicamente ogni cinque minuti, e allarmi visivi e acustici per avvertire i pazienti di episodi di ipoglicemia o iperglicemia. Tuttavia, sapere in anticipo se il glucosio sta per raggiungere livelli dannosi sarebbe molto più utile, poiché consentirebbe interventi proattivi.

La predizione della glicemia è un'area di ricerca consolidata nel campo del T1D, con una vasta letteratura che esplora diverse metodologie per migliorare la precisione delle predizioni. Sono stati proposti modelli di vario tipo, che vanno da approcci statistici tradizionali a tecniche avanzate di machine learning e deep learning. Nonostante questi progressi, lo sviluppo di algoritmi per la predizione della glicemia è ancora un problema aperto e la sua applicabilità clinica rimane una sfida continua a causa di fattori come la variabilità nelle risposte individuali, le imprecisioni nelle misurazioni dei sensori e la difficoltà nel catturare la complessità completa della dinamica del glucosio.

Una tendenza emergente nella predizione delle serie temporali è quella di trasformare i dati sequenziali in immagini, permettendo alle Reti Neurali Convoluzionali (CNN) di migliorare la precisione delle predizioni, grazie alle promettenti performance delle tecniche di visione artificiale. Questo approccio innovativo ha già mostrato prestazioni superiori in vari settori, come la predizione finanziaria, la classificazione dei modelli meteorologici e la diagnostica medica.

Questa tesi esplora l'applicazione della codifica delle serie temporali in immagini per la predizione della glicemia nel T1D. In particolare, i dati delle serie temporali della glicemia vengono trasformati in immagini attraverso tre tecniche di codifica: Recurrence Plot, Gramian Angular Field e Markov Transition Field. Queste rappresentazioni vengono quindi utilizzate come input per una Rete Neurale Convoluzionale (CNN) addestrata a prevedere i livelli di glicemia su vari orizzonti temporali. Per valutare questo approccio, viene condotta un'analisi comparativa di diverse lunghezze di input e orizzonti di predizione, confrontando una rete Long Short-Term Memory (LSTM) che elabora direttamente i dati CGM. Il dataset utilizzato in questa ricerca consiste nei dati CGM di 100 individui virtuali monitorati per 30 giorni, generati utilizzando il simulatore UVA/Padova Type 1 Diabetes Mellitus Simulator (T1DMS).

La tesi è organizzata come segue: il Capitolo 1 illustra una breve introduzione alle complicazioni, alla gestione e alle tecnologie relative al T1D, fornendo una panoramica degli algoritmi predittivi principali nello stato dell'arte.

Il Capitolo 2 presenta l'analisi del dataset utilizzato in questo lavoro per la valutazione delle metodologie di predizione.

Il Capitolo 3 fornisce una descrizione dettagliata di tre tecniche di codifica delle serie temporali in immagini esplorate in questo studio: Recurrence Plot (RP), Gramian Angular Field (GAF) e Markov Transition Field (MTF). Inoltre, vengono presentati gli algoritmi predittivi sviluppati per la predizione della glicemia: un modello autoregressivo, una rete Long Short-Term Memory (LSTM) e una rete neurale convoluzionale (CNN).

Il Capitolo 4 presenta i risultati, offrendo importanti approfondimenti sulle capacità predittive dei modelli CNN addestrati su immagini e confrontando le loro prestazioni con quelle del modello di riferimento LSTM addestrato su dati di serie temporali.

Il Capitolo 5 analizza i principali risultati ottenuti dallo studio, esplorando le implicazioni e confrontando le prestazioni dei modelli sviluppati.

Infine, il Capitolo 6 riassume le conclusioni dello studio, evidenziando alcune limitazioni e suggerendo promettenti direzioni per la ricerca futura

Contents

Abbreviations

AI	Artificial intelligence
AIC	Akaike Information Criterion
AP	Artificial pancreas
AR	AutoRegressive
ARIMA	AutoRegressive Integrated Moving Average
ARX	Autoregressive with Exogenous input
BG	Blood glucose
CDSS	Clinical decision support systems
CGM	Continuous glucose monitoring
CHO	Carbohydrate
COD	Coefficient of Determination
CNN	Convolutional Neural Network
CRNN	Convolutional Recurrent Neural Network
DKA	Diabetic ketoacidosis
GADF	Gramian Angular Difference Field
GAF	Gramian Angular Field
GASF	Gramian Angular Sum Field
HHS	Hyperosmolar hyperglycaemic state
LSTM	Long short-term memory
LVX	Latent Variable with Exogenous input
MARD	Mean Absolute Relative Difference
MDI	Multiple daily injections
MPC	Model Predictive of Control
ML	Machine learning

MTF	Markov Transition Field
NIR	Near-infrared
NN	Neural Network
PH	Prediction horizon
RMSE	Root Mean Square Error
RP	Recurrence Plot
ROC	Rate of glucose change
RF	Random Forest
SMBG	Self Monitoring Blood Glucose
SIPs	Smart insulin pens
SVR	Support Vector Regression
T1D	Type 1 diabetes
T2D	Type 2 diabetes
TD	Time Delay
WHO	World Health Organization

List of Figures

List of Tables

Chapter 1

Introduction

This chapter is structured as follows: Section 1.1 provides an overview of diabetes, discussing its main complications and the technologies employed for its management. Section 1.2 discusses blood glucose (BG) prediction, emphasizing its importance and analyzing the main linear and nonlinear algorithms. Section 1.3 briefly introduces a new prediction approach using time series-to-image encoding, highlighting its advantages and the major applications in the literature. Finally, Section 1.4 outlines the objective and structure of this thesis.

1.1 Diabetes: Overview of the disease and Technologies for its Management

Diabetes is a chronic, metabolic disease characterized by impaired insulin production and resulting in altered glucose homeostasis. This disease is one of the leading causes of blindness, kidney failure, heart attacks, and stroke, having a significant impact on an individual's daily life. Moreover, according to the World Health Organization (WHO), approximately 1.5 million deaths globally are attributable to diabetes each year, highlighting the urgent need for effective management and prevention strategies [1]-[2]. Diabetes can occur either as type 1 (T1D), in which the pancreas is unable to produce insulin, or as type 2 (T2D), characterized by insulin resistance and reduced insulin secretion. Less common, gestational diabetes mellitus (GDM) can develop during pregnancy due to hormonal changes that cause temporary insulin resistance, increasing the risk of complications both during pregnancy and at birth [3].

This thesis will focus on T1D, which is an autoimmune disease consisting of the destruction of pancreatic β-cells, which are responsible for insulin production. Without insulin, glucose levels accumulates in the bloodstream leading to acute complications, such as diabetic ketoacidosis and hyperglycemia (persistent BG concentration > 180 mg/dL) [4]. To compensate the lack of insulin, individuals with T1D have to administer exogenous insulin to keep glucose levels within the safe range of [70-180] mg/dL. However, due to the various factors, such as: physical activity, stress, under/overestimation of the amount of carbohydrates in a meal and overestimation of insulin dose, people with T1D can experience hypoglycemic episodes (BG concentration < 70 mg/dL) [5]-[6].

Hyperglycemia

Hyperglycemia is a condition that occurs when the body is unable to use insulin effectively, leading to the accumulation of glucose in the blood instead of using it as an energy source. Diabetic ketoacidosis (DKA) and hyperosmolar hyperglycemic state (HHS) are two severe and life-threatening complications of hyperglycemia, particularly common in T1D. DKA occurs when cells start burning fat to provide energy, resulting in the production of ketones that accumulate in the blood, leading to long term complications such as neuropathy, retinopathy, and micro-/macrovascular heart diseases. On the other hand, HHS causes extreme hyperglycemia without significant ketosis or acidosis, which often results in severe dehydration and altered mental status, including confusion, lethargy or even coma [4].

Hypoglycemia

Although most diabetic patients experience episodes of hyperglycemia more frequently, a small percentage faces a high incidence of severe hypoglycemias, with BG levels falling dangerously below the normal range. This condition can result from insulin overdose, alcohol abuse, inadequate carbohydrate (CHO) intake, or intense exercise. Hypoglycaemia can cause severe symptoms like confusion, excessive sweating, tremors, and if not treated promptly, can rapidly progress to acute cognitive dysfunctions, cardiac arrhythmias, coma or even death [7].

1.1.1 T1D Management Technologies

The standard T1D therapy aims at maintaining BG in the safe range during the day and it mainly consists in the administration of insulin, combined with dietary modifications, lifestyle adaptations, and frequent BG monitoring [8]. This section reviews the main technologies for T1D management, including glucose monitoring sensors, insulin delivery systems, and applications such as artificial pancreas and decision support systems.

Self-monitoring of Blood Glucose (SMBG)

The process of BG self-monitoring requires patients to independently monitor their glucose levels several times a day using glucometers and to calculate and adjust insulin doses according to various factors, including diet (particularly CHO intake), physical activity, and subjective factors like stress or illness. To take daily glucose measurements, patients have to prick their finger with a small lancet to collect a drop of capillary blood, which is then placed on a test strip and analyzed by a glucometer that displays the result on a digital screen within seconds (Fig. 1.1). Glucose readings are then recorded in a logbook or stored in the glucosemeter's electronic

memory so that patients can keep track of changes in glucose levels over time and adapt the treatment plan according to daily needs [9]. The frequency at which T1D individuals should monitor their BG level varies depending on several factors like the type of diabetes, clinical condition of the patient, diet, physical activity and insulin dosage. However most experts agree that insulin-treated patients should monitor their BG at least four times a day, most commonly when they are fasting, before meals and before sleeping [10]-[11].
However, this monitoring technique has revealed many critical aspects. First, a significant factor is patients' lack of motivation and commitment to regularly collect measurements, combined with their insufficient knowledge in accurately interpreting the results. In addition, the limited frequency of SMBG measurements is insufficient to capture the daily variability of glucose and detect critical episodes, such as postprandial hyperglycemia or hypoglycemia from insulin overdose [12]. For these reasons, SMBG has been largely replaced by a new monitoring technique known as Continuous Glucose Monitoring (CGM), which began to gain prominence in the early 21 century [11].

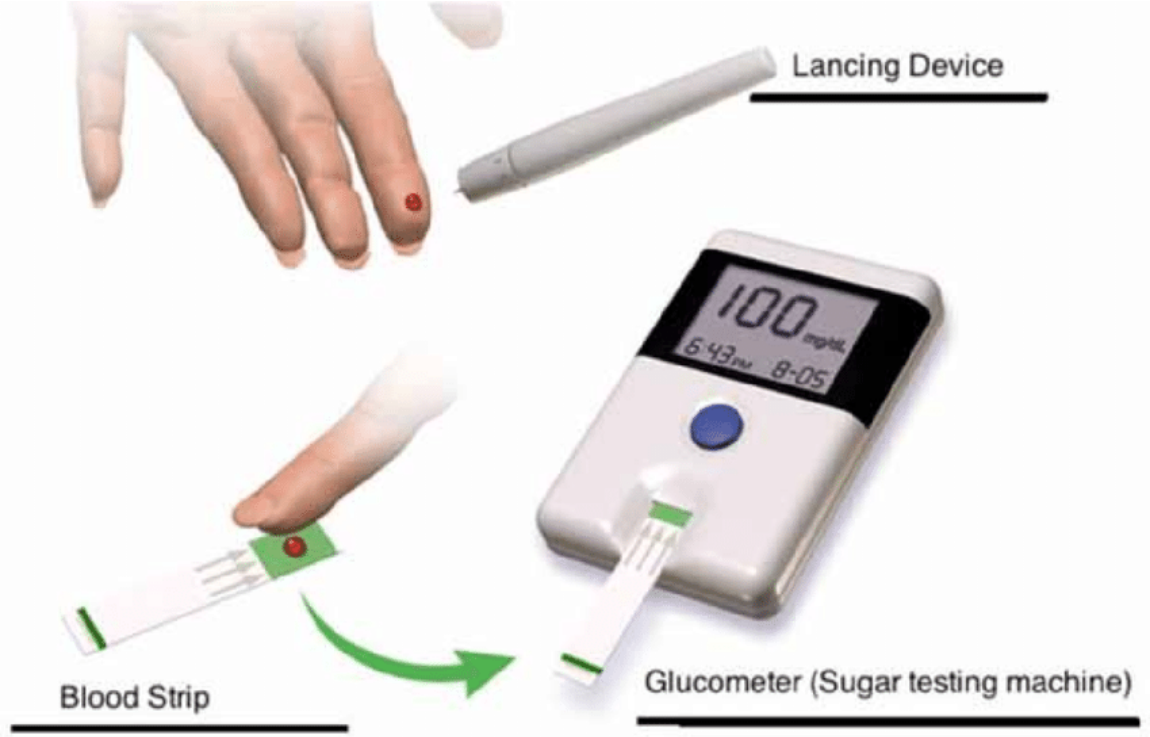


Figure 1.1: SMBG procedure: patient with T1D collects a blood sample from its finger using a lancet device, places the drop of blood onto a test strip, which is inserted into the glucometer. Source: `https://www.igi-global.com/gateway/article/full-text-html/284585`

Continuous glucose monitoring (CGM)

CGM is a technology that enables real-time tracking of BG levels throughout the day and night, providing patients with near-continuous measurements over several consecutive days (up to 10 days), usually at intervals of a few minutes, with a 5-min resolution. CGM devices allow not only continuous monitoring of glucose concentrations, but also provide information about the direction and rate of glucose changes (ROC). In addition, they are equipped with visual and acoustic alarms that alert users of the occurrence of hypo- and hyperglycemic episodes, enabling immediate interventions to prevent serious complications [13].

CGM devices consist of three main parts. The first is a small, minimally invasive needle sensor that is inserted into the subcutaneous tissue on the abdomen or the back of the arm to measure BG concentrations in the interstitial fluid. This sensor typically uses a platinum electrode doped with glucose oxidase, deposited on the needle, which catalyzes glucose oxidation. This results in the production of gluconolactone, hydrogen peroxide, and an electrical current signal, which is finally converted into a glucose concentration. The second part is composed of a transmitter that wirelessly sends the glucose data to the third component of the system, a receiver. The receiver consists of a software application stored on a smartphone, smartwatch, or other dedicated device that displays real-time measurements on a monitor, allowing the user to track their glucose levels continuously [14]-[15]. An example of a CGM device is shown in Fig.1.2.

Figure 1.2: Example of CGM device: Eversense® E3. On the right picture the three main components: transmitter, CGM sensor and receiver (smartphone or smartwatch). Source: `https://www.diabetes.ascensia.it/eversense-e3/`

The introduction of these sensors has revolutionized BG monitoring and opened new possibilities for daily diabetes management, addressing many of the limitations and challenges associated with traditional SMBG:

- it allows a more effective management of daily BG levels (and TD1 therapy as well), significantly improving the detection of hypo- and hyperglycemic events that SMBG might not capture, as can be seen in Fig.1.3 [15];
- it makes glucose monitoring less invasive and more convenient for patients, since it doesn't require frequent blood sampling (as in SMBG);
- the portability of these devices and the ability to receive an alert of the occurrence of extreme glycaemic events allow patient to become more self-sufficient in the management of both therapy and critical episodes. [13].

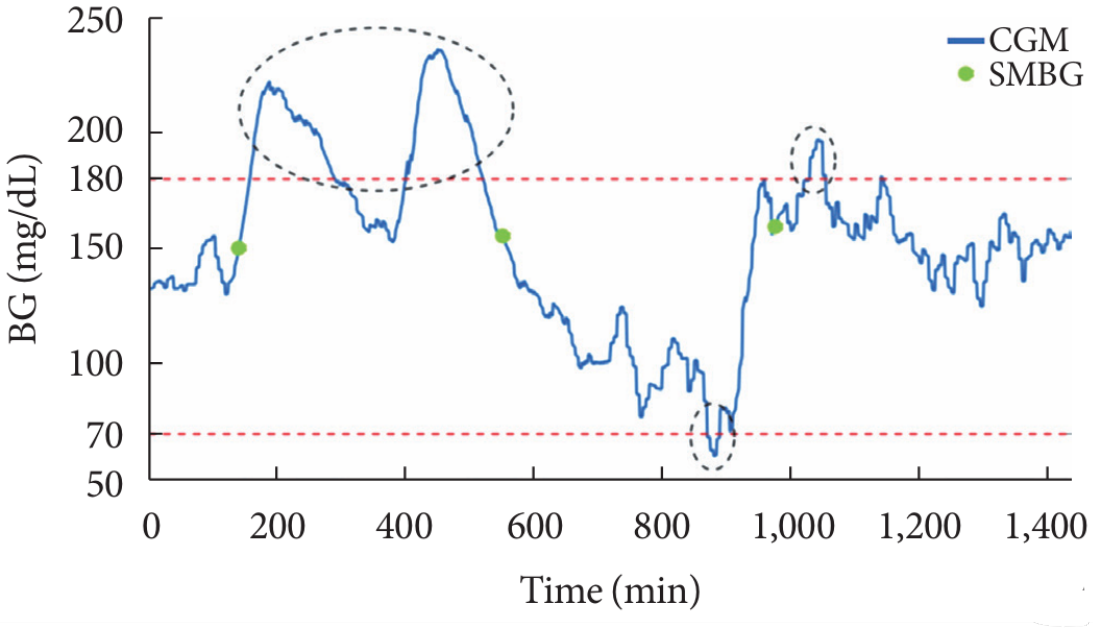
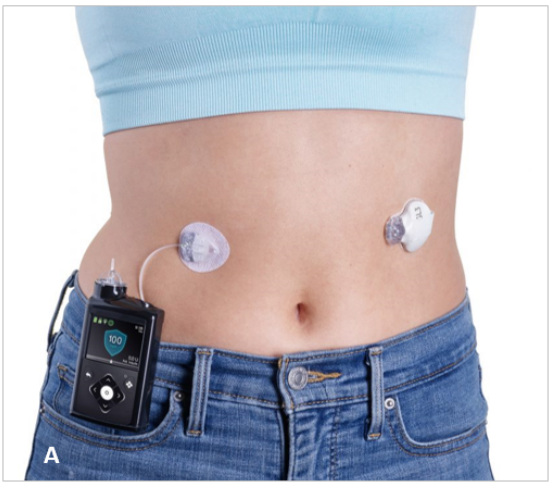


Figure 1.3: Representative BG monitoring data obtained with SMBG (in green) and CGM (in blue). The dashed circles denote hyperglycemic and hypoglycemic episodes which, using only SMBG measurements, are not detectable. Source: Cappon et al. [15]

Insulin pump and smart pens for insulin therapy

Portable Subcutaneous Continuous Insulin Infusion (CSII) pumps, more simply known as 'insulin pumps', are medical devices used to deliver insulin continuously through the skin. Unlike traditional insulin injections, which require multiple self-administration throughout the day,

pumps provide a constant flow of insulin on a customized schedule to mimic the pancreas' natural insulin secretion. Insulin administration occurs primarily in two ways: a continuous infusion of rapid-acting insulin throughout the day and night, known as basal insulin, and discrete doses of rapid-acting insulin, called insulin bolus, which are given by the user for meals or to correct elevated BG levels. The functioning of an insulin pump (Fig.1.4a) is based on the use of a catheter that is inserted under the skin (usually in the abdomen) and connected to an insulin reservoir [16]. The most recent pumps (Fig.1.4b), often known as 'patch' or 'pod' pumps, remove the need for traditional connection tubes and manual insertion of the infusion set catheter, simplifying the insulin delivery process [17]-[18]-[19].



(a) Traditional pump MiniMed™ 780G. Source: `https://www.medtronic-diabetes.com/it-IT`

(b) Tubeless pump Omnipod DASH®. Source: `https://www.omnipod.com/en-gb/diabetes-hub/pod-uni/insulin-pump`

Figure 1.4: Insulin delivery via insulin pump: an example of a traditional and a more innovative device.

Other widely used devices for insulin administration are insulin pens, which allow patients to select a precise dose of insulin and inject it with minimal preparation, making them especially useful for people who manage their diabetes independently. Smart insulin pens (SIPs), represent another step forward in diabetes management. These devices are similar to traditional insulin pens but offer enhanced functionality, in fact they not only administer insulin but they are also able to communicate to other devices via Bluetooth connectivity, record and store data of meal and insulin injections, as well as calculate the insulin bolus and trigger some reminder alerts (Fig.1.5). Moreover, SIPs allow more precise dosing, reduce pain at the injection site, and their ease of use makes them especially suitable for people with impaired vision or motor difficulties [20]-[21].

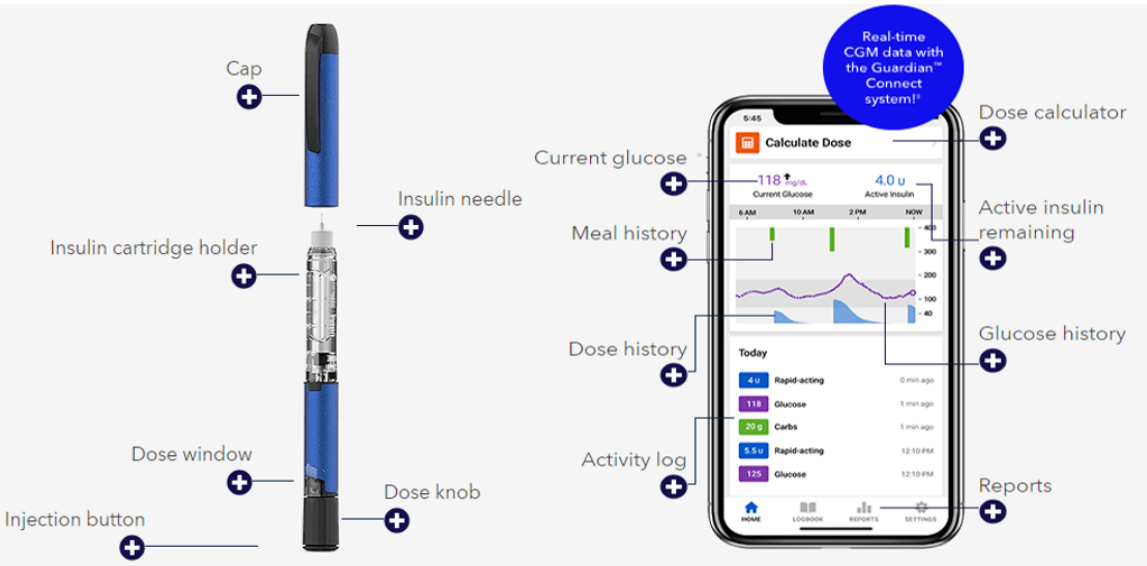


Figure 1.5: Insulin delivery via smart insulin pen. On the left, Medtronic InPen™ with its various components. On the right, a diabetes management system app that connects, via Bluetooth, to the SIP. Source: `https://origin.medtronicdiabetes.com/products/inpen-smart-insulin-pen-system`

Artificial Pancreas and Decision Support Systems

The use of CSII pump and CGM technologies has enabled the development of advanced methods for insulin delivery, such as artificial pancreas (AP)[22]. AP is an integrated technology that aims to reproduce the function of the human pancreas by allowing the automatic subcutaneous infusion of insulin on the basis of the BG concentration, measured continuously by the CGM sensor. AP systems consist of three main elements, shown in Fig.1.6, connected wireless:

1. a **subcutaneous sensor** for continuous glucose reading;
2. a **continuous insulin delivery device** (e.g. an insulin pump);
3. a **control algorithm** that can be integrated into the pump or associated with a separate unit (e.g. a smartphone) [23].

The algorithm is the key element in the functioning of the AP as it determines the amount of insulin to be infused subcutaneously into a patient based on the glucose value recorded by the sensor in the interstitial fluid. Several algorithms have been proposed over time, but two are commonly used: Proportional-Integral-Derivative (PID) and Model Predictive of Control (MPC). The PID algorithm is a feedback control method that continuously adjusts insulin delivery by responding to glucose levels in real time, with the aim of reducing deviations from the target glucose range. The MPC algorithm, on the other hand, is a model-based approach that predicts future glucose levels based on past data and predicted trends, optimising insulin delivery

proactively [24]. AP systems have been shown to provide better glycemic control than sensor-augmented pump therapy, with a lower risk of hypoglycemia in T1D patients [22].

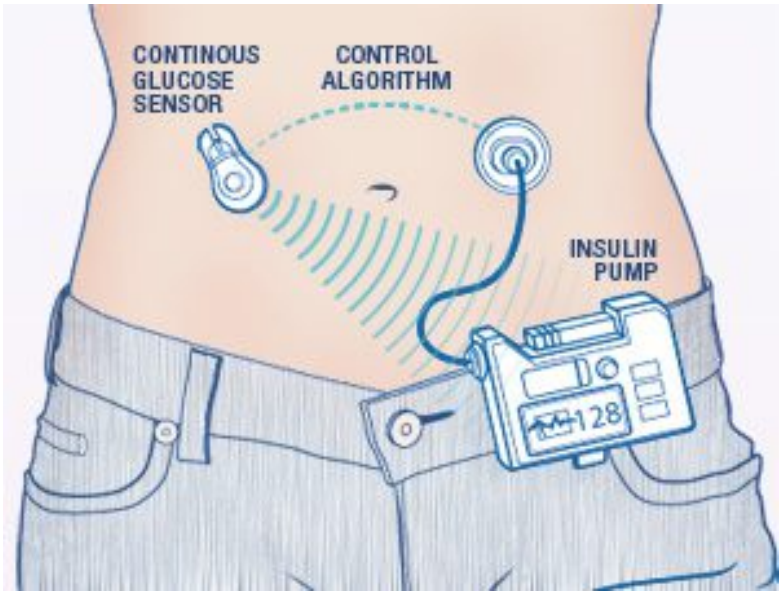


Figure 1.6: Artificial pancreas system with its three main components: continuous glucose sensor, insulin pump and control algorithm. Source: `https://integrateddiabetes.com/closed-loop-artificial-pancreas-system-update/`

More recently, the diabetes research comunity has intensively focused on the development of clinical decision support systems (CDSS). CDSS are defined as ”technology-based systems designed to improve healthcare delivery by enhancing medical decision-making with targeted clinical knowledge and up-to-date health information” [25]. In practical terms, they are advanced tools that employ algorithms and artificial intelligence (AI) models to analyze patient’s clinical data and medical parameters, providing personalized recommendations for diabetes management. For example, a recent review by Contreras et al. [26] showed that DSSs have been developed to:

1. compute the optimal insulin dose [27]-[28];
2. support carbohydrate content estimation in meals using computer vision methodologies [28]- [29];
3. detect unannounced meals [30], physical activity [31], and sensor faults [32];
4. detect adverse glycemic events to respond promptly to their effects [33]–[34]].

CDSS can be classified into various categories based on factors like intervention timing, and whether they have active or passive delivery. In particular, they are frequently classified as

knowledge-based or non-knowledge based systems, depending on how they process and deliver clinical recommendations [25]. Knowledge-based CDSS utilize a set of predetermined rules, typically formulated as IF-THEN statements, in order to produce an output or a targeted action and guide decision-making. These rules are established using evidence derived from literature-based, practice-based, or patient-directed evidence. In contrast, non-knowledge-based CDSS rely on AI, machine learning (ML), and statistical models to process decisions. These systems analyze large amounts of clinical data and identify patterns that can be useful for medical decisions, without adhering to static rules defined by experts. Both types of CDSS have three main common components with subtle differences (as can be seen in Fig .1.7) :

- **Base**: programmed rules (in the case of knowledge-based CDSS) or algorithms that process decisions based on AI and ML (for non-knowledge-based CDSS). Both systems also require access to clinical data.
- **Inference engine**: a component that applies the rules or algorithms to the patient's clinical data to generate an output or action, which is presented to the end user (e.g., physician).
- **Communication mechanism:** the website, application, or EHR frontend interface, through which the end user interacts with the system [25].

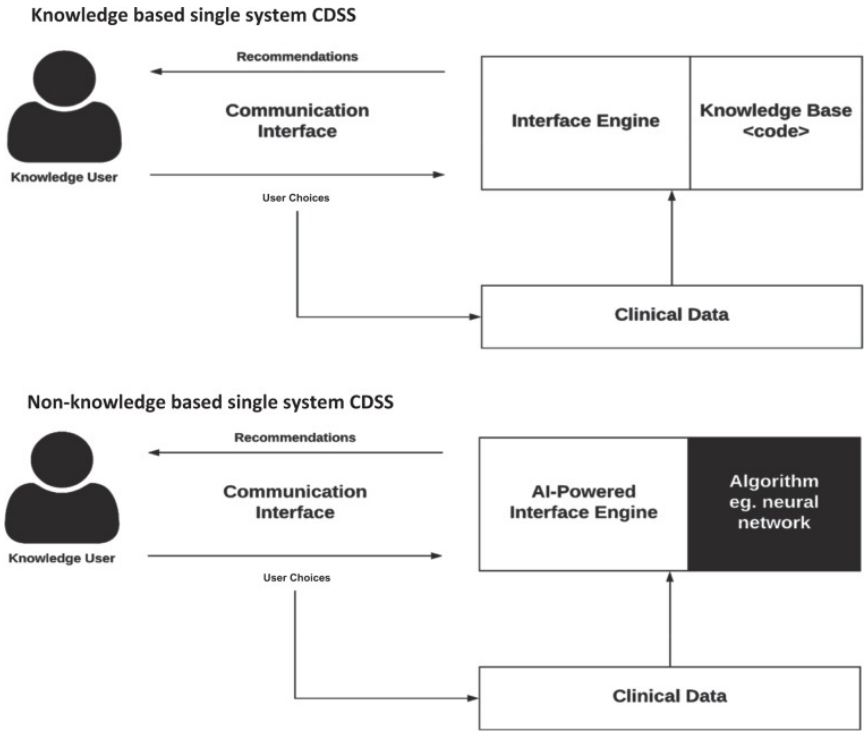


Figure 1.7: Diagram of key interactions in knowledge and non-knowledge based CDSS. Source: Sutton et al. [25].

1.2 Blood Glucose levels Forecasting

The daily diabetes therapy requires a number of actions that add a further burden to individuals with T1D. The development of CGM systems marked a significant advance in diabetes management; however, rather than simply responding to cases of hypo- or hyperglycemia after they have occurred, it is clearly more beneficial to predict ahead in time and to act before the occurrence of such episodes. For this reason, in recent years there has been an increasing interest in the field of BG forecasting, with the aim of developing more accurate and reliable models to anticipate variations in glucose levels [35]. This section will discussed the main techniques employed for BG forecasting, with a focus on linear and nonlinear models developed in literature.

1.2.1 The Role of BG Forecasting for improving T1D management

BG prediction aims to estimate the precise concentration of glucose levels several minutes into the future, as illustrated in Figure 1.8. From a patient perspective, knowing with a useful time anticipation if BG is approaching harmful levels could enable targeted proactive interventions, thus resulting in a reduced risk of acute hypoglycemic or hyperglycemic episodes and revolutionize T1D management.

Despite advancements in this area and a vast number of published contributions, the development of an accurate BG forecasting algorithm remains a challenging task because of the complex and highly individualized nature of glucose metabolism [36].

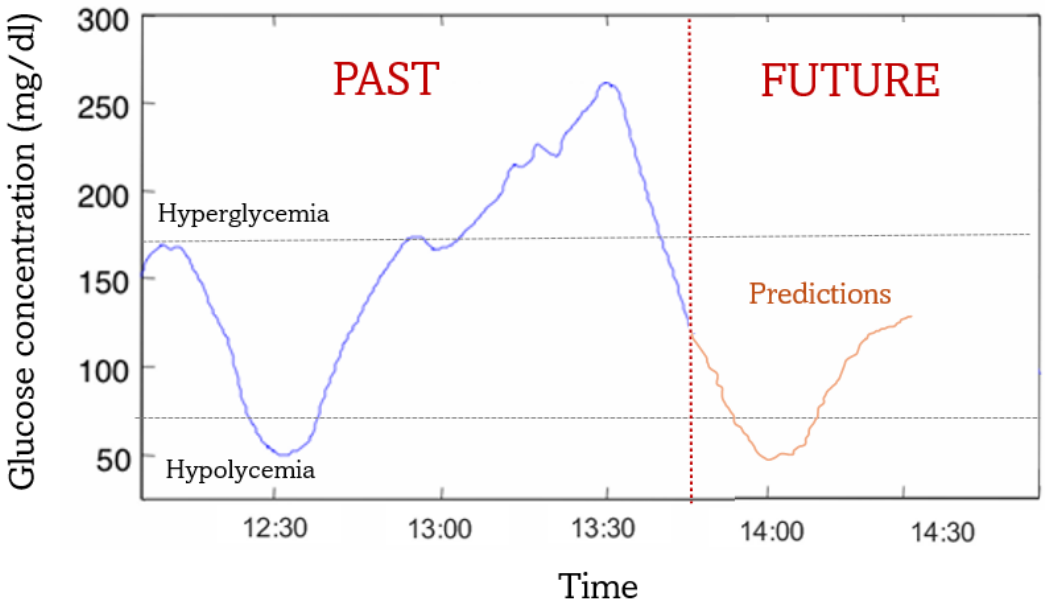


Figure 1.8: Example of a 30-minute ahead BG forecast.

Several factors can directly influence glucose levels, among which the history of BG values, insulin, physical activity, and dietary intakes are the prominent ones. Additionally, other conditions such as body mass index, stress level, hours of sleep, the presence of illnesses, medications, smoking habits, and alcohol consumption can also have an important impact [37]. An ideal BG predictor should incorporate as much information as possible to effectively track and predict BG levels, as shown in Fig. 1.9.

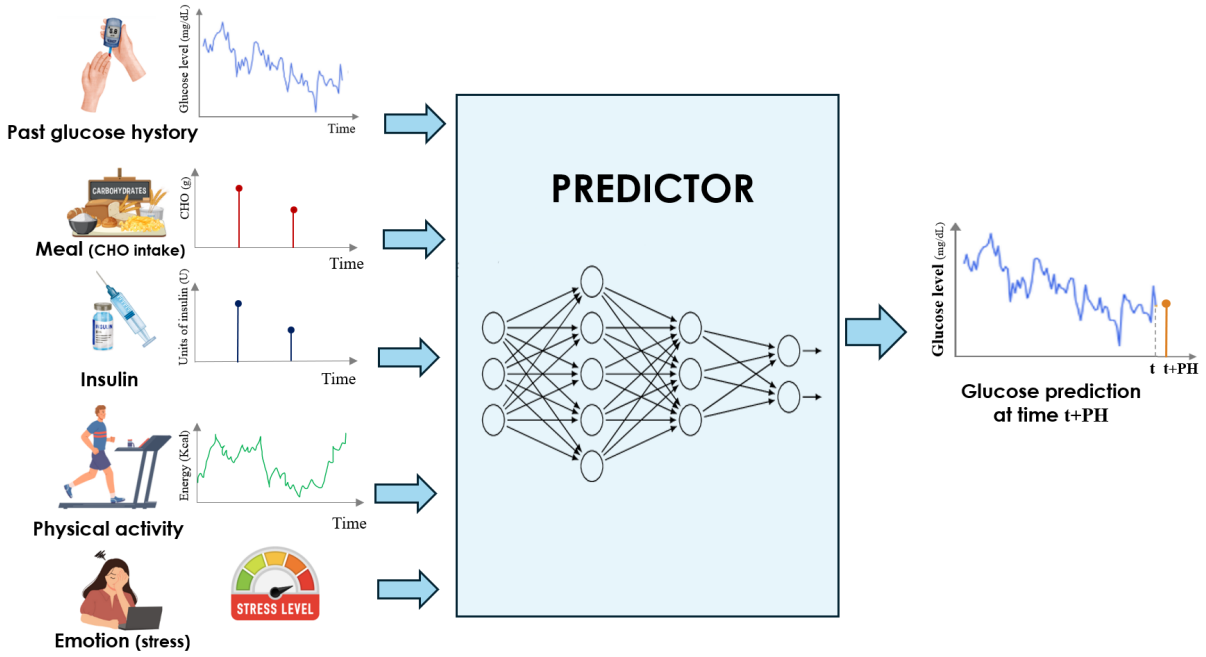


Figure 1.9: Schematic overview of an ideal glucose predictor able to exploit most of the inputs influencing glucose concentration.

Another essential aspect of glucose prediction is the prediction horizon (PH), i.e. the time interval in which the model is able to anticipate glucose levels. As the prediction horizon increases, for example moving from a 30-minute PH to a 60- or 90-minute PH, predictions will be subject to greater uncertainty. This is because, as one moves further ahead in time, the variables and factors that may influence BG levels, such as food intake, physical activity, hormonal changes and drug interactions, increase [38].

1.2.2 Overview of Blood Glucose Forecasting Algorithms

Over the years, several approaches for predicting glucose levels have been explored, ranging from traditional time series approaches such as autoregressive models to more complex machine and deep learning algorithms. This subsection will present an overview of the main linear and nonlinear approaches investigated in the literature for BG forecasting.

Linear Models

Autoregressive (AR) and autoregressive moving average (ARMA) models are among the first and most widely used linear approaches for BG levels forecasting. They rely on the assumption that future glucose values can be expressed as a linear combination of past glucose readings, capturing temporal dependencies in the data. Sparacino et al. [39] tested a first-order autoregressive model, which continuously adapts the model coefficient to predict future glucose concentrations. Results demonstrated that glucose can be predicted in advance with a 30-minute PH, a sufficient margin to mitigate hypoglycaemic events with sugar ingestion. Reifman et al. [40] developed a time invariant AR model, which parameters are personalized to patients data. This study showed that autoregressive models, when correctly adjusted to each patient's specific data, provide highly accurate and clinically acceptable short-term predictions of BG levels. Gani et al. [41], on the contrary, built a time invariant AR model of order 30, demonstrating that a single universal AR model can be used to make predictions on different subjects. Katsarou et al. [42] deal with an ARMA model for short-term glucose prediction of T1D patients. The optimal model parameters were determined through partial autocorrelation plots and the Akaike information criterion (AIC), and were identified as p=2 and q=2. Otoom et al. [43] employed an online estimation algorithm based on an ARIMA (AutoRegressive Integrated Moving Average) model, which continuously updates with observed historical data. Among the models considered, ARIMA$(3,0,0)$ demonstrates a good ability to provide acceptable estimates of BG levels. To improve the predictive accuracy, variants of AR and ARMA models that include external inputs, such as ARX and ARMAX, were employed. Unlike standard autoregressive models, which rely only on previous glucose values, these multivariate approaches also take into account the effect of other factors, like administered insulin doses, CHO intake, and physical activity. Finan et al. [44] proposed a third-order ARX model for glucose prediction considering CHO values and insulin doses. Such model was presented in two versions: one time-invariant and one time-varying, recursively identified at each step. Eren-Oruklu et al. [45] developed an ARMAX model considering the effect of a subject's physical activity and emotional stimuli such as stress on glucose metabolism.

Despite the wide use of linear models for BG forecasting, they present significant limitations due to their assumption of linearity and proportionality, which may be inadequate for capturing the complexity of glycemic dynamics. As a result, these models often struggle to provide accurate predictions in dynamic and complex situations, such as those influenced by external factors (i.e., meals, insulin administration, exercise and stress).

Non-Linear Models (Machine and Deep Leaning Techniques)

The advent machine learning and deep learning techniques, has revolutionized the approach of time series prediction. These methods are able to learn complex, non-linear relationships between variables, overcoming the limitations of traditional models. Machine learning includes a number of algorithms that allow systems to learn from data and improve over time without being explicitly programmed. In the context of BG prediction, these ML techniques can effectively capture complex patterns and dynamics of glucose fluctuations influenced by various factors. Among machine learning methods, Support Vector Regressor (SVR) and Random Forest (RF) have proven to be particularly effective for predicting glucose levels [46].
The task of glucose prediction in T1D patients was addressed in the context of SVR and RF by Georga et al. [47]-[48]. Both studies revealed that these two ML techniques are efficient for an accurate prediction in both the short and long term, especially when all available information is efficiently exploited.

Deep learning is a subset of machine learning that focuses on training algorithms to automatically extract complex patterns from large amounts of data. It relies on neural networks (NN), computational models inspired by the structure and function of neurons in the human brain, where multiple layers of artificial neurons, known as 'nodes', work together in deep neural networks to progressively transform and interpret data. Over the past decade, deep learning has led to significant advances in computer vision, disease diagnosis and healthcare, mainly due to its ability to automatically extract complex features and minimize the need to manually engineer features [49]. In this regard, many studies have been reported in the literature: Zhu et al. [50] proposed a convolutional neural network (CNN)-based model to predict the glucose level in 30 minute PH. The dataset used, OhioT1DM, includes data from 6 patients with T1D collected over an 8-week period, and it's the source of four input fields: glucose levels, insulin events, CHO intake and time index. Li et al. [51] introduced GluNet, a deep neural network (DNN) framework that exploits dilated convolutional layers to improve CNN performance. Results showed that GluNt achieved better glucose predictions in terms of RMSE, MARD, and in particular time lag, compared to rival models (NNPG, SVR, LVX, and ARX). Jaloli et al.[52] proposed a new prediction model for BG, built using a stacked architecture combining a CNN with a long short-term memory (LSTM). This study demonstrated the superiority of the CNN-LSTM model over ARX, SVR, LSTM and CNN-based models, attributing its higher performance to the integration of convolutional layers and LSTM in a layered structure that effectively captures complex and hidden features in multivariate datasets.

1.3 Time Series to Image Encoding for time series Forecasting

A successful field for the application of machine and deep learning algorithms is computer vision, where CNNs have demonstrated outstanding performance by extracting and identifying complex patterns from images. Inspired by these successes, researchers have begun exploring techniques to transform sequential data into image-like representations, thus allowing CNNs to capture temporal dependencies more effectively. This section provides a review of the main contributions dealing with time series to image encoding techniques [53].

1.3.1 State-of-art algorithms based on time-series encoding approaches

In literature the three main transformations employed to transform a time series into an image-like data are Recurrence Plots (RP), Gramian Angular Fields (GAF), and Markov Transition Fields (MTF). A more in-depth description of the techniques is available in Chapter 3, Section 3.2. There are only few contributions dealing with the use of time-series-to-image and they span different fields, from: financial forecasting to medical diagnosis.

For instance, Zhiguang et al. [54] proposed a framework for encoding time series data as images to enhance classification tasks using Convolutional Neural Networks (CNNs). Specifically, they focused on encoding time series into two types of representations: Gramian Angular Fields (GAF) and Markov Transition Fields (MTF). The task involves classifying time series data from 12 different datasets, applying Tiled Convolutional Neural Networks (Tiled CNNs) to the images with the aim of assessing whether these image-based representations can outperform the results of existing time series classification methods. This approach demonstrated competitive performance when compared to five state-of-the-art methods, achieving better classification accuracy for most datasets. Moreover, results showed that the combined GAF-MTF images led to the best results, outperforming individual GAF or MTF representations on ten out of twelve datasets.

Barra et al. [55] employed this approach to the financial field, developing a system for predicting the future trend of the U.S. stock market, based on a set of CNNs trained on GAF images generated from time series data. This method outperformed the Buy-and-Hold strategy and existing predictive models, enhancing robustness and reducing forecast variability. Therefore, this approach demonstrated the effectiveness of combining CNNs with time series-to-image encoding for financial trend prediction. Giannetti et al. [56] address the problem of identifying weather phenomena from the received signal level (RSL) by proposing a novel approach that maps time series into images through a transformation function. Again, CNN models were used for classification, and various data augmentation methods were integrated for both time series

and generated image data. Experimental results confirmed the validity of this approach, demonstrating the model ability to achieve high classification accuracy.

Ko et al.[57] applied this method to medicine for the first time, specifically to diagnose schizophrenic patients, by converting EEG signals into images. EEGs were converted to images using two time series image conversion algorithms, Recurrence Plot (RP) and Gramian Angular Field (GAF). The converted images were then learned by CNN models built on the basis of VGGNet. This study showed that the classification performance improved compared with previous studies. Specifically, between the two algorithms used for image conversion, the DL model learned through GAF showed significantly higher classification accuracy.

1.4 Aim and Structure of the Thesis

The use of time-series-to-image encoding approaches, combined with Convolutional neural networks (CNNs), has demonstrated promising performance across various fields. However, it has not yet been applied to the development of forecasting algorithms for T1D. To address this gap, this thesis aims to explore different time-series-to-image encoding techniques and develop deep learning models for BG forecasting. The central hypothesis is that transforming CGM time-series data into image-like representations enables deep neural networks to capture complex temporal patterns that traditional CGM-based models may overlook, ultimately improving predictive accuracy. To evaluate this approach, we conduct a comparative analysis of different input lengths and prediction horizons, benchmarking a long short-term memory (LSTM) network that directly processes CGM data against a deep CNN trained on the image-like transformations of the CGM sequences. An overview of the proposed predictive framework is illustrated in Fig. 1.10.

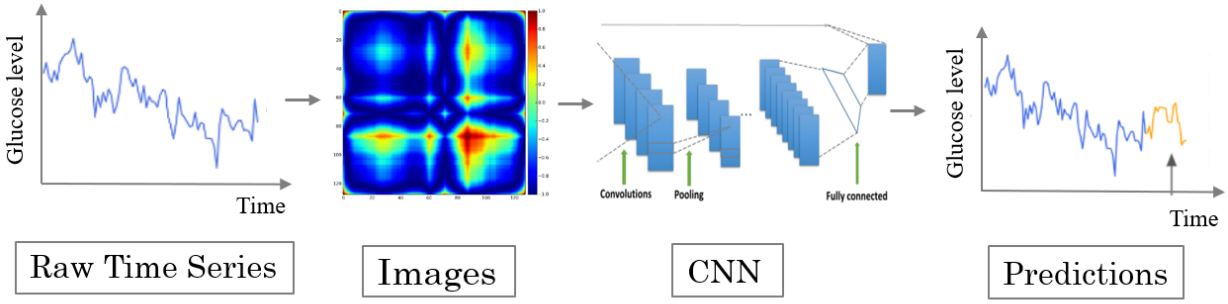


Figure 1.10: Conceptual framework of the project: time series-to-image encoding, development of a CNN model based on images, prediction of future glucose levels.

Structure of the thesis

- **Chapter 2** will present the description of the dataset used in this work for the assessment of the forecasting methodologies.
- C**hapter 3** provides a detailed explanation of three time-series-image encoding techniques investigated in our study: Recurrence Plot (RP), Gramian Angular Field (GAF), and Markov Transition Field (MTF). Moreover the chapter presents an overview of the predictive algorithms developed for future BG forecasting : an Autoregressive (AR) model and a Long Short-Term Memory (LSTM) network that that directly process CGM data, and a Convolutional Neural Network (CNN) trained on the images obtained from our dataset.
- **Chapter 4** details the results obtained from this study, providing important insights into the predictive capabilities of CNN models trained on images, as well as a comparison with the predictive performance of the LSTM model trained on time series data.
- **Chapter 5** analyzes the main findings obtained from the study, exploring their implications and comparing the performance of the models developed.
- **Chapter 6** summarizes the conclusions of the study, highlighting some limitations and outlining promising directions for future research

Chapter 2

Dataset

This chapter provides an overview of strategies adopted to obtain predictions of BG levels in T1D using the time-series-image encoding approach previously described. Section 2.1 will detail the dataset used, generated through the UVA/Padua Type 1 Diabetes Mellitus Simulator (T1DMS), and the use of Automated Glucose Data Analysis (AGATA) software for glucose time series analysis. Section 2.2 will illustrate the techniques employed to transform BG time series into images. In particular, three methodologies will be described: Recurrence Plot (RP), Gramian Angular Field (GAF) and Markov Transition Field (MTF). The predictive models developed for BG prediction are presented in section 2.3: a linear model based on Autoregression (AR), a Long Short-Term Memory (LSTM) network, and a Convolutional Neural Network (CNN). The first two models were trained using glycemic time series, while the third leveraging the three types of images generated as model inputs. Finally, section 2.4 will describe the criteria and metrics for evaluating the predictive performance of the models employed.

2.1 UVA/Padua Type 1 Diabetes Mellitus Simulator (T1DMS)

The dataset employed in this thesis for developing BG forecasting algorithm consists of 100 virtual individuals monitored for 30 days. The dataset was generated using the UVA/Padova Type 1 Diabetes Mellitus Simulator (T1DMS). The T1DMS is a large-scale maximal computer model which includes 13 differential equations and about 36 subject specific parameters that allow to model glucose, insulin and glucagon dynamics in T1D population. It was developed with the primary purpose of improving diabetes research, allowing to simulate and test treatment protocols identical to proposed clinical studies. Pre-clinical experiments conducted with the UVA/Padova T1DMS are designed at the level of individual subjects, which provides insights into intra- and interpersonal differences observable in humans. By modeling real-life conditions, including meal size, insulin dosing and variability in insulin sensitivity throughout the day, T1DMS provides a highly realistic representation of BG dynamics in diabetic patients [58].The key advantage of generating a simulated dataset is the ability to control and adjust critical factors such as insulin dosage, meal timing, and other patient-specific variables. This enables the generation of consistent and reproducible data, allowing prediction models to be tested under carefully controlled conditions that may be difficult or impractical to replicate in real clinical settings [59].

More specifically, the dataset used for this thesis contains simulated data for 100 T1D subjects, recorded at 5-minute intervals over a period of one month. Each subject's data includes three main features: BG concentration (mg/dL), amount of insulin administered (IU), and car-

bohydrate intake (g CHO)In Table 2.1, a summary of the main descriptive statistics for the three features is presented, providing a general overview of the dataset.

	Blood Glucose [mg/dL]	Insulin Bolus [UI]	Meal [g CHO]
Mean value	149.95	0.539	0.142
Median value	148.25	0.48	0.14
Standard deviation	17.98	0.22	0.01
Global minimum	40	0	0
Global maximum	399	472.29	62.6

Table 2.1: Descriptive Statistics for Blood Glucose, Insulin Bolus, and Carbohydrate intake.

2.2 Dataset Analysis with AGATA

For the analysis of the dataset, the Automated Glucose Data Analysis (AGATA) toolbox [60] was used. AGATA is a software designed for the exploration of glucose data, offering a series of functionalities for pre-processing and analysis of CGM data. Specifically, it allows the visualization of glucose profiles, computation of glucose control metrics, detection of adverse events, and assessment of predictive-algorithms performance. AGATA's versatility and ease of use make it a valuable tool for both clinical research and the development of innovative diabetes management strategies [61].

In Fig. 2.1 and 2.2 the global and daily glucose profiles for a representative subject from the dataset employed in this thesis are shown. From the global glycemic profile it can be observed a higher frequency and longer duration of hyperglycemic episodes compared to hypoglycemic ones, which is a typical pattern observed in T1D patients. The analysis of the daily glycemic profile reveals that hyperglycemic peaks are almost corresponding to the times of the three main meals, highlighting the increase of glucose levels due to CHO intake.

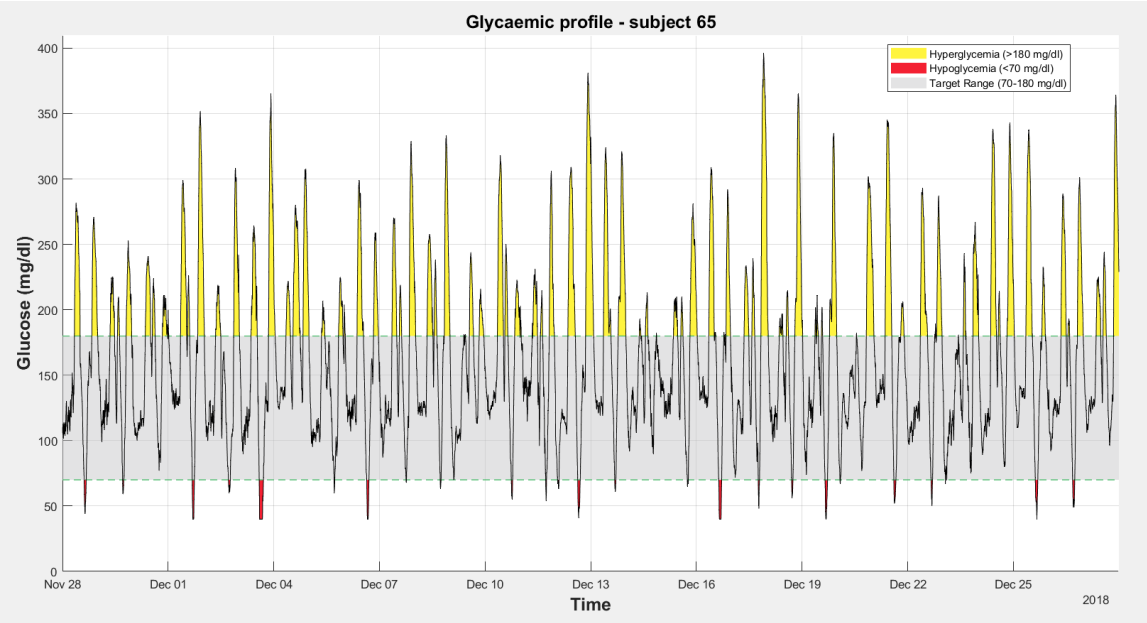


Figure 2.1: Global glycemic profiles of a representative subject from the simulated dataset. The black continuous line indicates CGM trace over the whole recording period, hyperglycemic events are highlighted with yellow areas, and hypoglycemic ones with red areas.

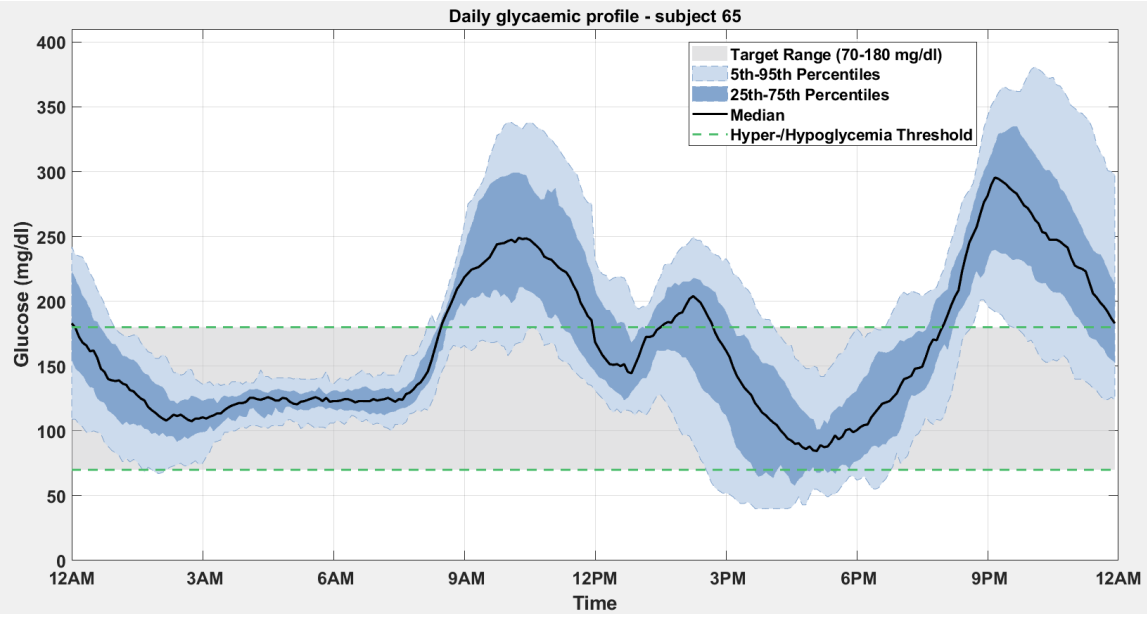


Figure 2.2: Daily Glycemic profiles of a representative subject from the simulated dataset. The median daily profile is represented by a bold black line, while the dark blue and light blue shaded areas highlight the 25th to 75th and 5th to 95th percentiles, respectively.

In Fig. 2.3, the histogram of BG rate of change (ROC) for the same representative subject is depicted. In the context of blood glucose, the ROC is defined as the rate at which glucose levels change over time, and it is calculated as the difference between the BG value at two consecutive time points.The ROC can be expressed as follows:

$$\text{ROC} = \frac{BG(t + \Delta t) - BG(t)}{\Delta t}$$

Where:

- $BG(t)$ is the glucose level at time t,
- $BG(t + \Delta t)$ is the glucose level at a subsequent time $t + \Delta t$,
- Δt is the time interval between the two points.

The histogram in Figure 2.3 exhibits a bell-shaped curve, indicating a normal distribution of ROC glucose data, with most variations concentrated near zero. This distribution is further confirmed by the red curve overlapping the histogram, which represents the respective fitted Gaussian distribution. The symmetry of the distribution is also evident, indicating a balance in the frequency of upward and downward fluctuations in glucose levels, suggesting that there is no pronounced trend toward sudden increases or decreases.

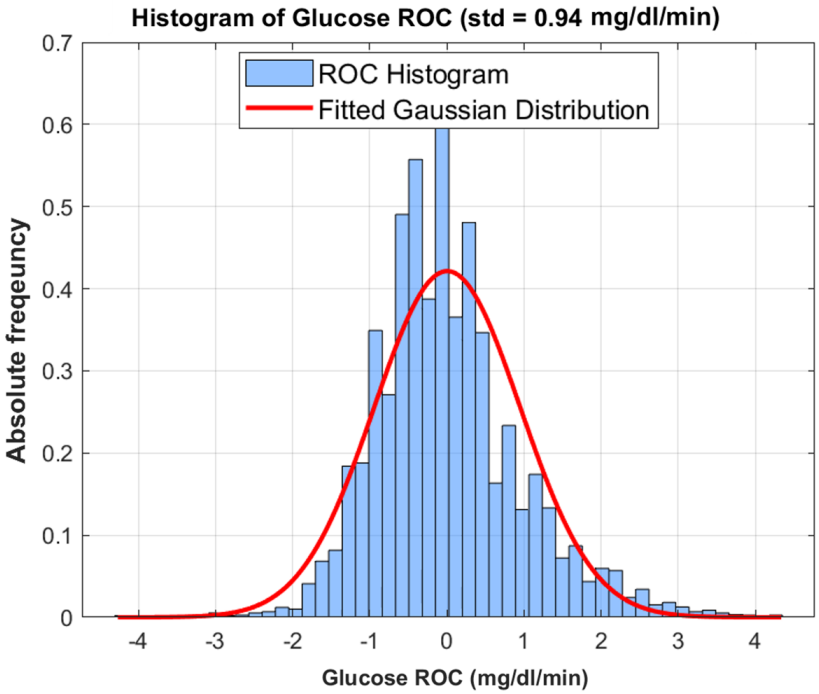
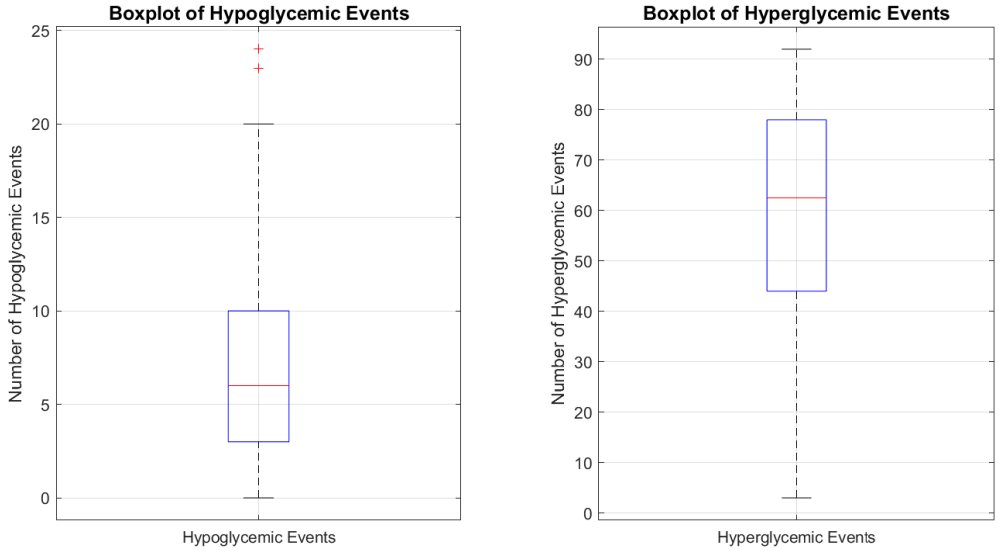
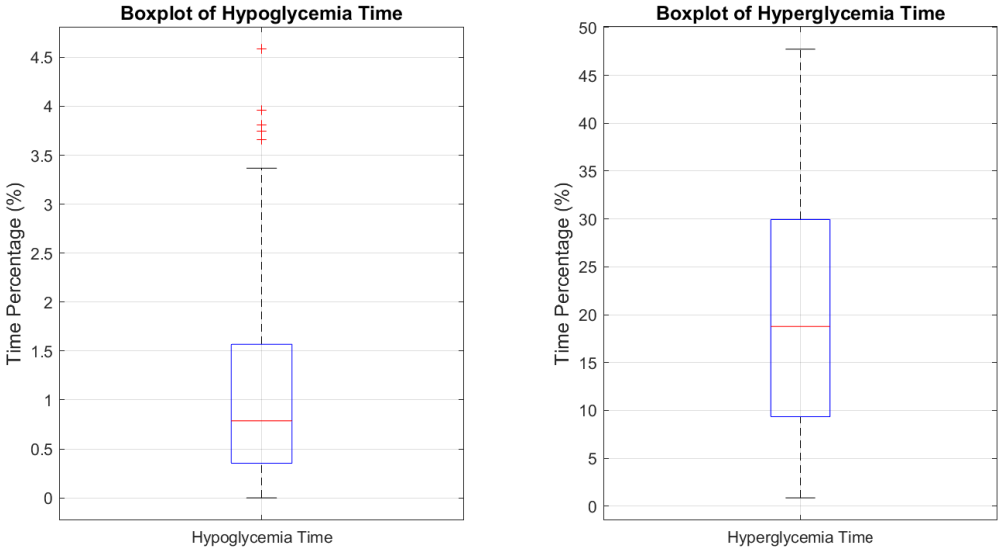


Figure 2.3: Histogram of glucose ROC (in blue) and the respective fitted Gaussian distribution (in red) for a representative subject of the simulated dataset.

For a more in-depth examination of critical events, both the number of hypo- and hyperglycemic episodes and the percentage of time spent in these conditions during the recording period were investigated. The results of this analysis are shown in Fig. 2.4, which illustrates the boxplots for the number of critical episodes (Fig. 2.4a) and the percentage of time spent in these conditions (Fig. 2.4b).



(a) Boxplot of the number of critical glycemic events.



(b) Boxplot of the time spent in critical glycemic states.

Figure 2.4: Critical glycemic events analysis: a comparison of hyperglycemic and hypoglycemic episodes.

Both types of boxplots reveal a marked difference between hypoglycemic and hyperglycemic events. Regarding the number of episodes, it can be observed that hypoglycemias are significantly less frequent than hyperglycemias, with a median of 6 and 62.5, respectively. This indicates that, on average, subjects experience a much higher number of hyperglycemic events than hypoglycemic ones. One can also observe that some subjects do not experience any hypoglycemic episodes during the observation period, as evidenced by the minimum value of zero in the left boxplot. The same pattern is reflected in the boxplot of time spent in critical conditions: the median time spent in hypoglycemia is only 0.79% of total monitoring time, compared with a median time spent in hyperglycemia of 18.77%. This suggests that hyperglycemic events are not only more frequent but also have a longer duration than hypoglycemic episodes, reflecting the greater prevalence and persistence of hyperglycemia as a critical condition in the T1D population. Finally, some outliers are observed for both the number of hypoglycemic events (4 outliers) and the time spent in hypoglycemia (5 outliers). Among these, two individuals deserve special attention, subjects 60 and 65, who are outliers in both categories. This means that they not only experience significantly more hypoglycemic episodes than average (23 and 24 events, respectively), but also spend a particularly high percentage of time in hypoglycemia (3.75% and 3.8% of the monitoring period, respectively).

Risk metrics were then analyzed for subjects who have recorded the highest number of hypoglycemic and hyperglycemic episodes (subjects 60, 65, 14 and 92) . The following risk metrics were analyzed:

- **LBGI** (Low Blood Glucose Index): measures the average BG level during hypoglycemic periods. A high LBGI indicates a higher frequency of hypoglycemic episodes.
- **HBGI** (High Blood Glucose Index): measures the average BG level during hyperglycemic periods. A high HBGI indicates a higher frequency of hyperglycemic episodes.
- **BGRI** (Blood Glucose Risk Index): an overall indicator of glycemic risk. A high BGRI suggests a higher risk of glycemic problems (hypo- and hyperglycemic events).
- **ADRR** (Average Daily Risk Range): measures the daily variability of BG levels, indicating fluctuations in BG throughout the day. A high ADRR suggests greater fluctuation in daily glucose levels.

Based on the reported metrics, it can be observed that, regarding the LBGI index the subject with the highest values are 60 and 65, indicating a higher frequency of hypoglycemic episodes compared to the other subjects analyzed. Regarding the HBGI index, subjects 14 and 65 show a particularly high value, indicating less effective control during periods of hyperglycemia. Concerning the ADRR, subject 65 shows the highest value, suggesting significant daily variability

in glycemic levels and unstable glycemic control. Overall, the analysis of all risk metrics identifies subject 65 as the individual most prone to critical events, including both hypoglycemia and hyperglycemia. This highlights significant glycemic instability and a higher likelihood of critical episodes compared to the other subjects analyzed. From this analysis, the following risk metrics emerged in Table 2.2:

Risk Metrics	subject 60	subject 65	subject 14	subject 92
LBGI	1.1055*	0.9724	0.0148	0.2370
HBGI	3.7448	8.1697	8.9465*	6.1935
BGRI	4.8504	9.1421*	8.9613	6.4305
ADRR	39.7675	54.8690*	38.6245	36.9866

Table 2.2: Risk metrics for patients with the highest number of hypoglycemic and hyperglycemic episodes. Asterisks (*) indicate the highest values recorded for each risk index.

Chapter 3

Methods

This chapter provides a description of three time-series-to-image encoding techniques explored in our study: Recurrence Plot, Gramian Angular Field and Markov Transition Field (Section 3.1). Additionally, Section 3.2 presents an overview of three predictive algorithms employed for BG forecasting: an Autoregressive (AR) model and a Long Short-Term Memory (LSTM) network,and a Convolutional Neural Network (CNN) Finally, in Section 3.3, the evaluation metrics adopted to assess the performance of the our predictive models will be presented.

3.1 Image Encoding Techniques

In literature there are three main time-series-to-image encoding approaches: Recurrence Plot (RP), Gramian Angular Field (GAF), and Markov Transition Field (MTF). The following section provides a detailed explanation of these techniques, showing some examples of images generated from our dataset by varying the length of glycemic sequence.

3.1.1 Recurrence Plot (RP)

RP technique allows to convert a time series into a matrix where each point represents the Euclidean distance for each pair of values in the time series. Finally, the resulting matrix can also be binarized using a specific threshold. The procedure for building a RP from a time series is the following:

1. Consider a time series $x_1, x_2, \ldots, x_N$, where x_i is the value of the series at time t_i.

2. Calculate a distance measure between each pair of points in the time series. Typically, the Euclidean distance is used, which is defined as:

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^2}$$

   where x_i and x_j are values of the time series at time indices i and j, respectively.

3. Update the corresponding value in the matrix at position (i, j) with the calculated distance $d(x_i, x_j)$. This results in a complete matrix where each entry represents the distance between two points in the time series.

4. If one wants to binarize the matrix, compare the distance $d(x_i, x_j)$ to a pre-defined threshold and modify the matrix as follows:

- If $d(x_i, x_j) \leq \varepsilon$, set the value in the matrix at position (i, j) to 1 (typically represented as black).
- If $d(x_i, x_j) > \varepsilon$, set the value in the matrix at position (i, j) to 0 (typically represented as white).

The result is a matrix, where:

- White areas represent non-recurrent points (i.e., values in the time series that exhibit significant changes compared to other points).
- Black areas represent recurrent points (i.e., values that follow similar patterns over time).

Of note, the obtained RP matrix is typically symmetric because $d(x_i, x_j) = d(x_j, x_i)$ [62].

Example of RP images generated from glucose time series

Fig. 3.1 illustrates the time series and the corresponding RP matrices of the same subject for sequences of increasing length: 12, 144 and 288 samples, representing time periods of 1 hour, 12 hours, and 24 hours, respectively.

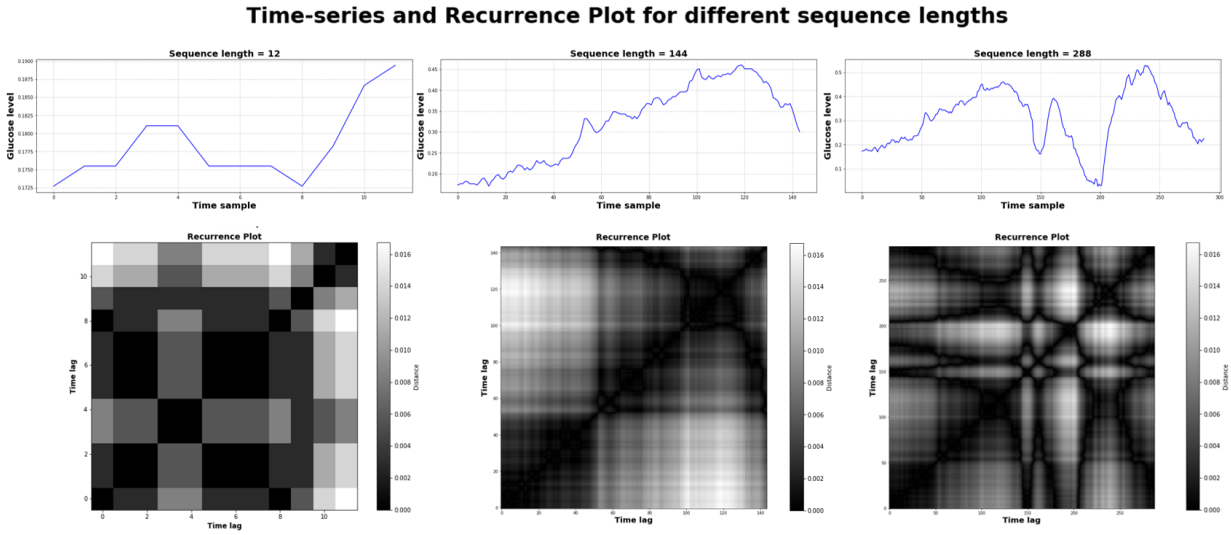


Figure 3.1: Comparison of time-series and RP matrices for different sequence length for the first sequence of the same subject of the training set. Letf panel: sequence of 12 samples (corresponding to a period of 1 hour), central panel: sequence of 144 samples (corresponding to 12 hours), right panel: sequence of 288 samples, corresponding to 24 hours.

The image on the left in Fig. 3.1 (corresponding to a sequence of 12 samples) shows some distinct black squares and rectangles in the RP matrix. These patterns indicate groups of points in which the distances between values in the glycemic series are particularly small, i.e., where

the time series shows a high degree of similarity. In other words, these structures highlight repetitive behaviors and recurrent temporal patterns within the time series data. From the upper left to the lower right corner it is possible to identify a sequence of black rectangles. These indicate segments of the glucose time series corresponding to an increasing and decreasing trend.

Increasing the length of the analyzed glucose signal from 1 hour to 12 hours (144 samples, central panel in Figure 3.1), the RP matrix offers a more comprehensive and detailed representation of the glycemic time series. The emergence of two black squares along the main diagonal reflects extended periods of similarity and recurrent behavior, with the first spanning approximately 50 samples and the second around 90 samples, capturing more intricate temporal patterns in glucose dynamics.

Finally, the image on the right panel (288 samples) reveals even more complex patterns, reflecting long-term recurrent trends in the time series. These structures are the result of the combination of repetitions at longer intervals, highlighting a more complex dynamic behavior of the glycemic data.

As a final remark, increasing the sequence length, i.e., considering a longer time history of the data, an increasing degree of detail in glycemic dynamics is observed from RP images. In fact, with longer sequences, more complex recurring behaviors and structures emerge in the RP matrix, allowing to capture phenomena that would not be visible with shorter sequences.

3.1.2 Gramian Angular Field (GAF)

GAF is a technique that converts a time series into a matrix, where each value indicates the angular variation between each pair of points in the time series. The procedure for constructing a GAF matrix (summarized in Fig. 3.2) is as follows:

1. Let $x_1, x_2, \ldots, x_N$ be a time series, where x_i is the value of the series at time t_i.

2. Normalize the time series by rescaling the values to the range $[0, 1]$, using the following transformation:
$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$
where $\min(x)$ and $\max(x)$ are the minimum and maximum values of the time series.

3. Convert the normalized values into polar coordinates by applying the following transformation:
$$\theta_i = \arccos(x'_i)$$

4. The elements of the Gramian matrix can be finally achieved by computing the cosine or sine of the sum or difference of the polar coordinates for each pair of points, obtaining

the **Gramian Angular Sum Field (GASF)** or the **Gramian Angular Difference Field (GADF)** respectively [62].

$$\text{GASFi,j} = \cos(\theta_i + \theta_j)$$

$$\text{GADFi,j} = \sin(\theta_i - \theta_j)$$

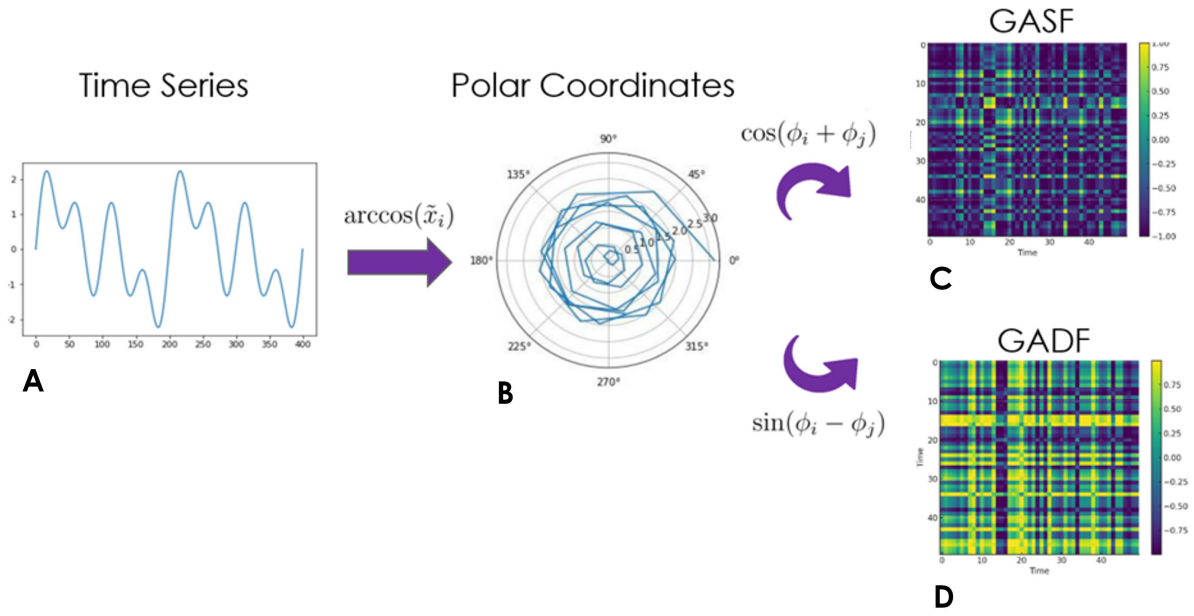


Figure 3.2: GAF image generation process. The time series in panel A is converted to polar coordinates in panel B. Then computing the cosine or sine of the sum or difference of the polar coordinates for each pair of points results in the GASF (panel C) and GADF (panel D) images, respectively

Example of GASF images generated from glucose time series

In Figure 3.3, the time series and the corresponding GASF and GADF matrices are shown for the same subject across sequences of increasing length: 12 samples (in the left panel), 144 samples (central panel), and 288 samples (right panel).

From these matrices, we observe colored structures ranging from yellow (corresponding to the maximum sum or difference of angles) to blue (indicating the minimum values of sum or difference of angles). For the shortest time series (12 samples, left panel), the GASF matrix highlights a yellow square on the main diagonal at sample 8, corresponding to a negative peak where the angular variation is most pronounced. In the GADF image, a yellow vertical stripe appears toward the end of the matrix (from sample 9 to 12), reflecting an increase in the glycemic

signal. This occurs because, at the slopes of the time series, the angular values associated with consecutive points vary rapidly, leading to large angular difference values.

As the size of the time series increases, i.e. moving from 12 to 288 samples (in the right panel of Figure 3.3), the GAF matrices become more detailed, capturing a richer representation of the glycemic signal dynamics. In fact, yellow areas appear in the GASF matrix, corresponding to the peaks (positive or negative) of the time series. In the GADF matrix yellow stripes are visible in a more defined way, indicating rises and slopes of the time series.

Therefore, similar to RP, GAF matrices demonstrate that longer sequences are better at capturing the temporal dynamics and patterns of glycemic data, providing a more comprehensive view of short- and long-term behaviors.

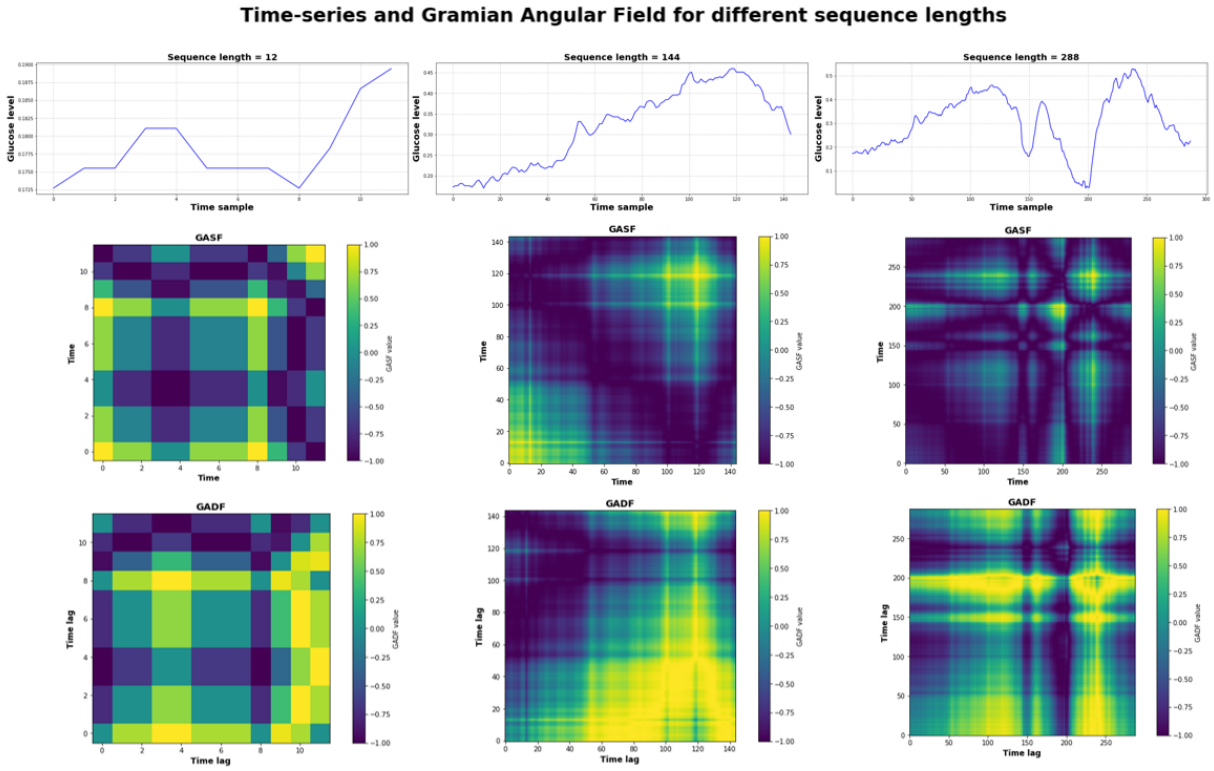


Figure 3.3: Comparison of time-series and GAF matrices (GASF on the top and GADF on the bottom) for different sequence length for the first sequence of a representative subject of the training set.

3.1.3 Markov Transition Field (MTF)

The last encoding approach investigated in this work is MTF, which is a technique used to transform a time series into a matrix where each element represents the transition probability between each pair of states in a discretized time series. Therefore this type of image provides a visualization of the way in which the values of the time series change from one state to another over time. The procedure for building a MTF matrix (summarized in Fig. 3.4) is the following:

1. Consider $x_1, x_2, \ldots, x_N$ a time series, where x_i is the value of the series at time t_i.

2. Discretize the time series by segmenting the values into intervals and assign a discrete state to each value x'_i.

3. Calculate the transition probability between each pair of states, defined as the frequency of the time series transitioning from state s_i to state s_j.

$$P(s_i \to s_j) = \frac{N_{ij}}{N_i}$$

where:

- N_{ij} is the number of transitions observed from state s_i to state s_j.
- N_i is the total number of occurrences of state s_i in the time series [62].

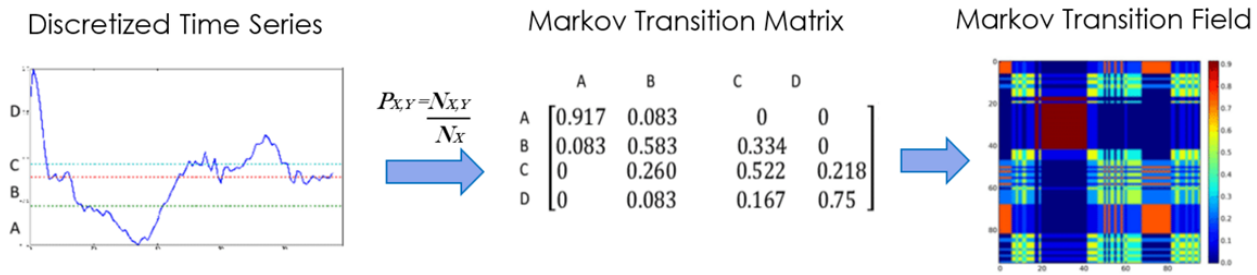


Figure 3.4: MTF image generation process. The time series is discretized into different states (panel A); the matrix containing the transition probabilities between the states is calculated (panel B); the matrix is converted into an MTF image (panel C)

Example of MTF images generated from glucose time series

As for the other encoding techniques, we provide an example of MTF transformations for different input lengths, which can be seen in Figure 3.5. As a parameter related to the number of states, a value of 15 was chosen, in line with parameters selected in the work of [63]. This choice enables a distinction between the different states of the time series, allowing the dynamics of the glycemic signal to be captured effectively. The MTF matrices show colored pixels, with yellow dots representing the highest transition probabilities and purple dots indicating the lowest, which respectively reflect areas where the time series is more or less likely to move from one state to another.

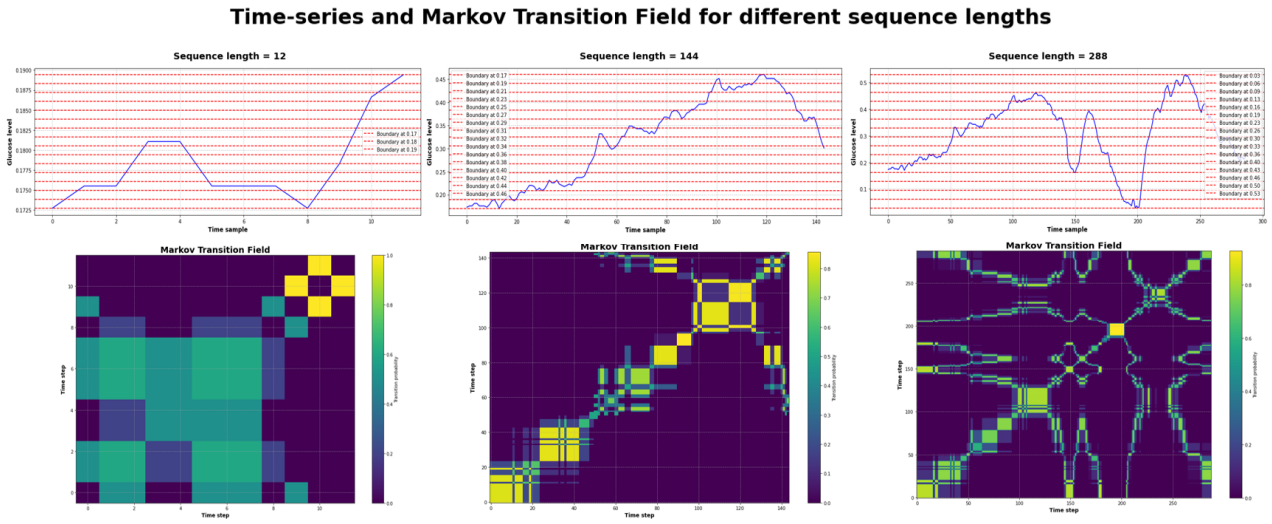


Figure 3.5: Comparison of time-series and MTF for different sequence length for the first sequence of a representative subject of the training set.

In the matrix representing a sequence of 12 samples (left panel of Figure 3.5), overlapping squares and rectangles appear, with yellow squares along the main diagonal from sample 8 to 12. This reflects an upward trend in the glycemic series, where the probability of state transitions increases due to the rapid rise in glucose levels.

In the central panel of Fig. 3.5, corresponding to a sequence of 144 samples, yellow areas appear along the main diagonal due to multiple rising and falling phases in the glycemic series, which increase the probability of state transitions. Notably, around the major peak (samples 90 to 130), yellow structures appear outside the main diagonal, indicating a consistent transition pattern before and after the peak.

Finally, for the 288-sample sequence (right panel), the MTF matrix reveals more complex structures, with yellow squares along the diagonal, corresponding to the rising and falling phases of the time series. In addition, curved lines emerge across the areas of the matrix related to the main peaks (positive and negative), suggesting a symmetrical trend with respect to these peaks.

3.2 Blood Glucose Prediction Models

This section provides an overview of the three BG forecasting methodologies adopted in this thesis. Specifically, a linear approach based on Autoregression (AR) and a nonlinear one based on a Long Short-Term Memory (LSTM) were developed from glucose time series data. The performance of these models were compared with the those obtained from a third model, a Convolutional Neural Network (CNN), trained on data transformed into three image types: RPs, GAFs, and MTFs For each of these models, predictions were calculated by varying the past history of the signal and considering five PH values: 30 minutes, 45 minutes, 60 minutes, 75 minutes, and 90 minutes. This choice allows both short- and long-term predictive accuracy to be assessed, providing a more comprehensive view of model performance.

3.2.1 Linear Model (AR)

The first approach employed for BG forecasting is the Autoregressive (AR) model, a statistical linear method that relies on historical data to predict future glucose levels. The AR model assumes that the future values of a time series can be expressed as a linear combination of its past values, plus an error term. The general form of an AR model of order p is:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t \tag{3.1}$$

where:

- X_t represents the value of the time series at time t,
- $\phi_1, \phi_2, \ldots, \phi_p$ are the coefficients estimated from the data,
- $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ are the past values of the time series,
- ϵ_t is the error term, which is generally regarded as white noise, i.e. , normally distributed with zero mean and constant variance [64]-[65].

For the predictions with the AR model, the following procedure was employed:

1. The dataset was divided into a 80% of training set (used to train the model on historical data) and a 20% of test set (used to evaluate performance on unseen data).
2. A Standard Scaler was applied to normalize glucose levels in both training and test sets.
3. The AR model was trained on historical data, capturing short-term dependencies in glucose data.

4. The model order, i.e, time lag, was chosen by applying a Random Search that tested values in the range from 1 to 20. The root mean square error (RMSE) value was used as selection criterion, which resulted in the best model order being 10.

5. Finally, for all test set data, BG prediction values were calculated for all selected PHs.

3.2.2 Long Short-Term Memory Neural Network (LSTM)

LSTM model is a type of recurrent neural network (RNN) specifically designed to address the limitations of traditional RNNs in handling long-term dependencies. LSTMs achieve this through their unique architecture, which includes gates that control the flow of information: a forgetting gate, an input gate, and an output gate. The forgetting gate determines which information should be discarded from the cell's memory, the input gate decides which new information should be added to the memory, and the output gate controls which part of the memory is passed to the next layer of the model. This mechanism allows the model to store only relevant information while filtering out irrelevant data, making it particularly effective for time series forecasting tasks. An illustration of the LSTM cell structure is shown in Fig. 3.6, which highlights its internal mechanisms.

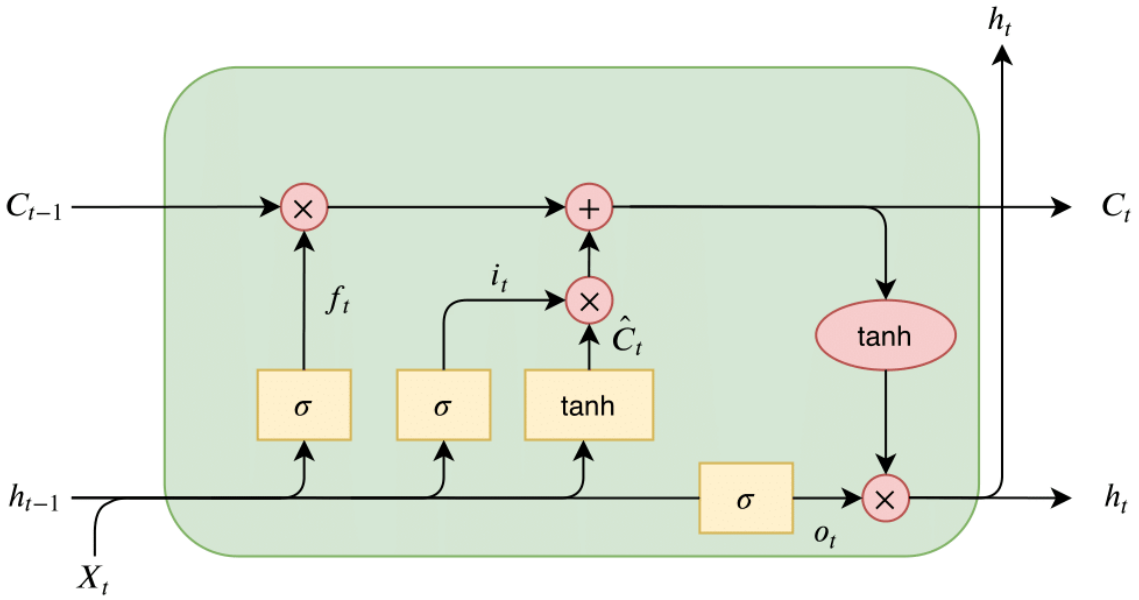


Figure 3.6: LSTM diagram with the forgetting gate (f_t), input gate (i_t) and output gate (o_t), and their respective activation functions (sigmoid and tanh). x_t is the input and h_t is the output of a cell. Source: Senshen et al. [66].

The LSTM model was chosen among nonlinear ones because it is particularly suited to capture both short-term and long-term temporal dependencies in glucose time series. In addition, its architecture allows it to effectively handle the sequential nature of time series data, making it an

ideal choice for this task [67].
In the literature, there are several scientific contributions that have employed LSTM models to predict glucose levels. For example, Aliberti et al. [68] proposed a multi-patient data-driven approach that used an LSTM networks to predict glucose levels based on a heterogeneous cohort of patients. Rabby et al. [69] developed an approach for glucose forecasting using a stacked LSTM-based deep recurrent neural network (RNN) combined with Kalman smoothing for sensor error correction. In the study conducted by Idrissi et al. [70] , the authors aim to evaluate and compare different Multi-Step-ahead Forecasting (MSF) strategies for predicting BG levels using LSTM Neural Networks. The more advanced long short-term memory (LSTM) networks, designed to capture long-term dependencies, were utilized by Prendin et al. in [71], Mosquera-Lopez et al. in [72], and Nasir et al. in [73].

For the predictions with the LSTM model, the following approach was employed:

1. The dataset was split into 60% of training set (used to train the model on known samples), 20% of validation set (allocated to tune hyperparameters), and 20% test set (to evaluate the model's ability to generalize to unseen data).

2. Overlapping temporal sequences were created to capture short-term dependencies in time series data.

3. The training set was further divided into a training set and a validation set, allocating 80% of the data for training and 20% for validation.

4. An LSTM model was developed.

5. During training, the model was evaluated on the validation set to monitor performance and avoid overfitting, with the root mean square error (RMSE) used as the evaluation metric.

6. Finally, for all test set data, BG predictions were computed for all selected PHs.

Multiple vs Single-Horizon strategy

BG forecasting can be addressed using two distinct approaches based on time horizon management. The Single-Horizon (SH) strategy involves the development of a separate model to predict BG values for each PH. In contrast, the Multi-Horizon (MH) strategy relies on a single CNN model that is able to predict glycemic values for multiple PHs in a single run. The model is designed to make continuous predictions over a time interval, such as from 5 to 90 minutes, with 5-minute intervals. This results in significantly lower computational costs with the MH

approach, particularly during training, where the loss function is minimized based on forecast errors across multiple horizons simultaneously rather than independently for each PH-specific model. In the literature, it is well established that MH-based forecasting tends to perform better than the SH approach. This is mainly due to the fact that the model is trained to make continuous predictions at close intervals, allowing it to better capture the temporal dynamics of glucose changes [74]. For these reasons, we focused on the MH strategy in our study, proving that it effectively outperforms the SH approach both in terms of prediction performance and computational efficiency.

3.2.3 The LSTM model Architecture employed in this study

For the development of the LSTM model, different architectures were tested, varying the number of layers, the number of neurons per layer, adding different normalization techniques such as batch normalization, dropout layers, and testing various regularizers to avoid overfitting. After conducting numerous experiments with these configurations, the most performing architecture resulted to be the simplest one, shown in Fig. 3.7. This architecture consists of a single LSTM layer with 5 units, followed by a dense layer that generates the final predictions for all PHs, ranging from 5 to 90 minutes. Regarding the selection of hyperparameters Adam was chosen as optimizer with a learning rate of 0.001 and ReLU was selected as activation function. A batch size of 256 was used, and the model was trained for 100 epochs with early stopping applied, using a patience of 10 epochs to prevent overfitting.

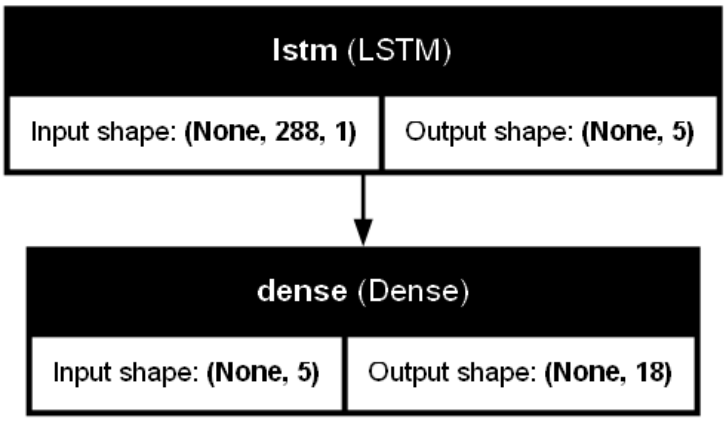


Figure 3.7: Architecture of the LSTM model adopted in our study.

3.2.4 Convolutional Neural Network (CNN)

The third approach employed for predicting BG levels involves the use of Convolutional Neural Networks (CNNs). CNNs are a class of deep learning neural networks commonly used in computer vision to process and analyze visual data, particularly for tasks such as object recognition, image segmentation, and feature extraction [75].
A typical CNN consists of multiple layers, including:

- Convolutional layers that apply filters (or kernels) to the input image to detect and capture local patterns such as edges, corners, and textures, producing a feature map.
- Pooling layers that resample the image to reduce its dimensionality and computational load while preserving the most significant features. These layers can be of different types, such as max pooling (which takes the maximum value of a certain area of the feature map), or average pooling (which calculates the average value).
- Fully connected layers in which the model learns the high-level combinations of features extracted from the filters to perform classifications or predictions [76].

A representation of a typical CNN architecture is shown in Figure 3.8, illustrating the processes of convolution, pooling, the flatten layer, and the fully connected layer.

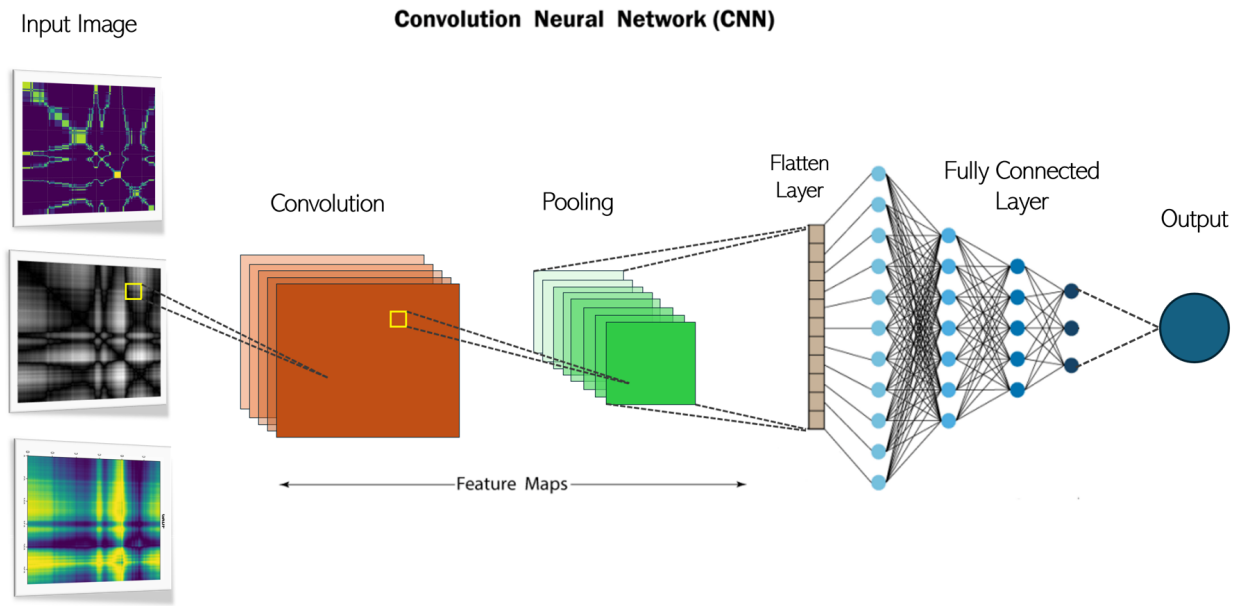


Figure 3.8: Architecture of a typical CNN model.

CNNs are potentially useful for forecasting when the input data is represented as images, as they can capture complex temporal patterns that may be more difficult to identify using only time series data. This is the reason why CNNs were chosen in our study to process three types of generated images (RP, GAF, and MTF) in order to obtain future glucose predictions.

3.2.5 The CNN model Architecture employed in this study

To perform glycemic prediction with CNNs, different architectures were tested for each image transformation method. For GASF, RP, and MTF, the same architecture was adopted, consisting of three convolutional layers with an increasing number of filters (32, 64, and 128) and kernels of size 3×3. The model also includes a fully connected layer with 32 neurons and ELU activation, followed by a Dropout layer (rate = 0.02) to reduce the risk of overfitting. An additional dense layer with 32 neurons and Sigmoid activation precedes the output layer. Fig. 3.9 illustrates a schematic of this architecture. For GADF, a similar architecture was employed, but with a reduced number of filters (2, 8, and 16) in the convolutional layers. All models were optimized using Adam with a learning rate of 0.0001. A batch size of 256 was selected, and the models were trained for 100 epochs with early stopping applied, using a patience of 10 epochs to prevent overfitting.

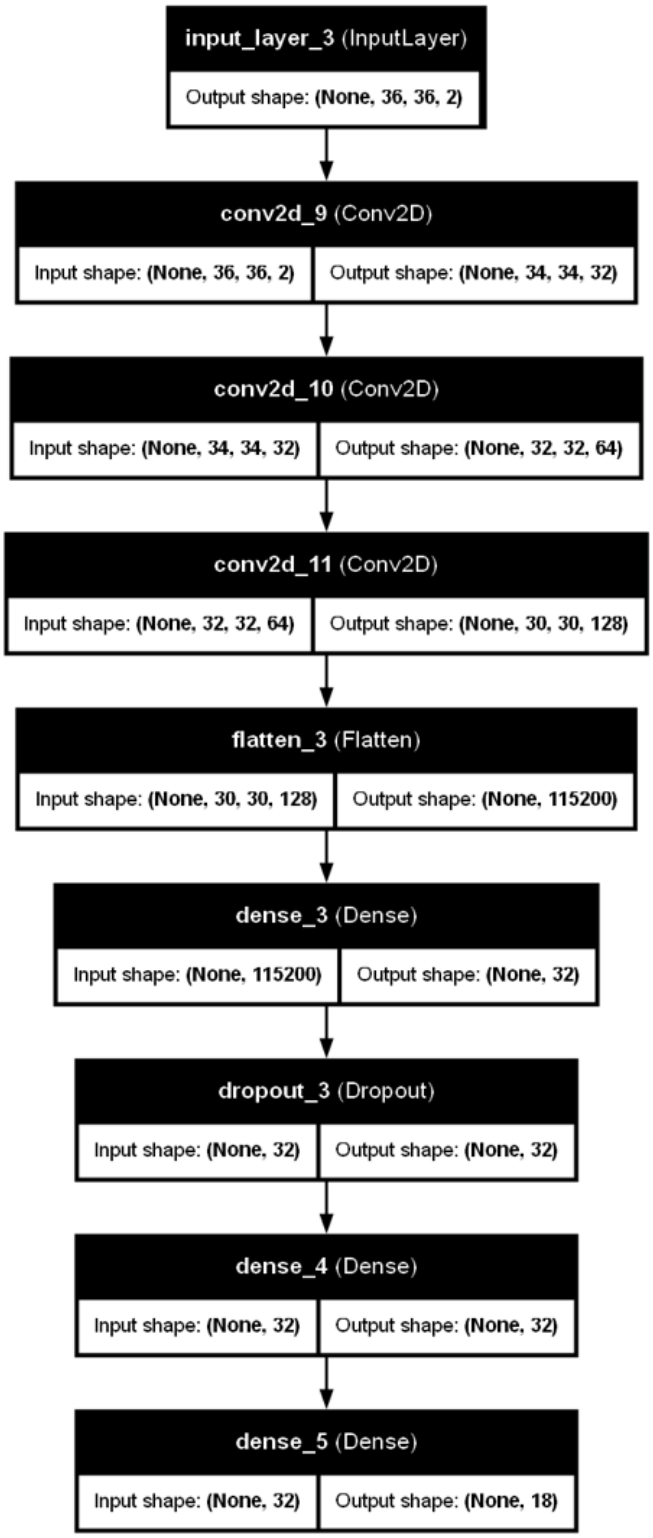


Figure 3.9: Architecture of the CNN model for GASF, RP and MTF techniques adopted in our study.

3.3 Criteria and Metrics for the Assessment

This section present the performance metrics adopted to evaluate the accuracy of the predicted profiles. To this purpose, four evaluation metrics were considered: the Root Mean Square Error (RMSE), the Mean Absolute Relative Difference (MARD), the Coefficient of Determination (COD), and the Time Delay.

3.3.1 Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is an error metrics that measures the discrepancy between the predicted glucose values and the actual measured CGM data. The RMSE is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y(t+PH) - \hat{y}(t+PH \mid t)\right)^2} \tag{3.2}$$

where:

- PH is the prediction horizon,
- N is the length of glucose data in the test set,
- $y(t+PH)$ is the actual glucose value at time $t+PH$.
- $\hat{y}(t+PH \mid t)$ is the predicted glucose value at time $t+PH$ based on data up to time t.

RMSE assumes only positive values, where RMSE $= 0$ corresponds to a perfect prediction. Lower values of RMSE indicate more accurate predictions, while increasing values of RMSE denote higher prediction errors, reflecting a greater discrepancy between predicted and actual glucose values.

3.3.2 Mean Absolute Relative Difference (MARD)

The second error metric computed is the Mean Absolute Relative Difference (MARD), which quantifies the average percentage error between the predicted glucose values and the actual measured CGM data. The MARD is calculated as follows:

$$\text{MARD} = 100 \times \left(\frac{1}{N}\sum_{t=1}^{N}\left|\frac{y(t+PH) - \hat{y}(t+PH|t)}{y(t+PH)}\right|\right) \tag{3.3}$$

, where:

- PH is the prediction horizon,
- N is the length of glucose data in the test set,
- $y(t + PH)$ is the actual glucose value at time $t + PH$,
- $\hat{y}(t + PH \mid t)$ is the predicted glucose value at time $t + PH$ based on data up to time t.

MARD expresses the average relative difference as a percentage, where lower values indicate higher prediction accuracy, i.e., predictions closer to actual glucose measurements, while higher values reflect larger relative errors.

3.3.3 Coefficient of Determination (COD)

Furthermore, we also considered the Coefficient of Determination (COD), indicating the percentage of variance in actual glucose values that can be explained by model predictions. The COD is defined as follows:

$$\text{COD} = 100 \times \left(1 - \frac{\|y(t + PH) - \hat{y}(t + PH \mid t)\|_2^2}{\|y(t + PH) - \bar{y}(t + PH)\|_2^2}\right) \tag{3.4}$$

where:

- PH is the prediction horizon,
- N is the length of glucose data in the test set,
- $y(t + PH)$ is the actual glucose value at time $t + PH$,
- $\hat{y}(t + PH \mid t)$ is the predicted glucose value at time $t + PH$ based on data up to time t.
- $\bar{y}(t + PH)$ is the mean of the actual glucose values at time $t + PH$.
- $\|x(t)\|_2$ is the Euclidean norm of the signal $x(t)$, computed as:

$$\|x(t)\|_2 = \sqrt{\sum_{t=1}^{N} (x(t))^2}. \tag{3.5}$$

The COD assumes its maximum value (i.e., 100%) if the glycemic profile predicted by the model exactly matches the target CGM signal. If the variance of the prediction error is equal to the variance of the signal, the COD becomes zero, indicating that the model is not able to account for the variability of the actual data. Therefore, the higher the COD value, the greater the model's ability to explain the variance in observed values, reflecting a more accurate prediction. There

is no lower bound for COD values, as they may also be negative.

3.3.4 Time Delay (TD)

Finally, the time delay (TD) is a metric that quantifies the time shift needed to maximize the correlation between predicted glucose values and measured data. Basically, it is calculated by determining the time shift that aligns the predicted and actual glucose traces as closely as possible. The time delay is defined as follows:

$$\text{Time delay} = \arg \min_{j \in [0, PH]} \frac{1}{N} \sum_{t=1}^{N-PH} \left(\hat{y}((t + PH \mid t) + j) - y(t + PH)\right)^2 \tag{3.6}$$

where:

- PH is the prediction horizon,
- N is the length of glucose data in the test set,
- $y(t + PH)$ is the actual glucose value at time $t + PH$,
- $\hat{y}(t + PH \mid t + j)$ is the predicted glucose value at time $t + PH$ based on the data up to time $t + j$,
- j is the time shift within the prediction horizon that minimizes the prediction error.

A time delay of zero suggests that the predictions are perfectly aligned with the observed values in time, while a non-zero value indicates that the predictions are ahead the actual measurements over time. A larger time delay leads to less accurate predictions, as it indicates a greater misalignment between the predicted and actual values.

Chapter 4

Results

This chapter presents the performance of the BG forecasting algorithms on the UVA/Padova Simulated Dataset. In particular, the evaluation metrics described in Section 3.3 will be reported for each prediction model implemented, considering different values of prediction horizon: 30, 45, 60, 75, and 90. Section 4.1 presents the results obtained for the Autoregressive model, Section 4.2 the performance of the LSTM model, and Section 4.3 the performance of the CNN model for all the time-series-to-image encoding techniques. In addition, a comparison of the performance for the three conversion techniques (RP, GAF and MTF) will be provided. Finally, Section 4.4 will compare the results of the LSTM model with those of the CNN model, obtained using the most performing conversion technique .

4.1 Linear Model (AR)

This section presents the prediction results achieved by the AR model. Table 4.1 shows the test set performance evaluated for five PHs: 30 min, 45 min, 60 min, 75 min, and 90 min. Performance metrics are presented as the Mean ± Standard Deviation across the test set population, which consists of 20 subjects.

Evaluation Metrics	**Prediction Horizon**				
	30 min	45 min	60 min	75 min	90 min
RMSE [mg/dL]	**16.73 ± 2.78**	23.01 ± 4.79	28.44 ± 6.70	32.99 ± 8.36	36.72 ± 9.73
MARD [%]	**8.11 ± 0.51**	10.81 ± 0.89	13.25 ± 1.34	15.48 ± 1.85	17.51 ± 2.33
COD [%]	**83.11 ± 7.34**	69.39 ± 10.74	54.62 ± 13.07	40.18 ± 14.33	26.92 ± 14.84
Time Delay [min]	**21.75 ± 2.45**	33.25 ± 3.73	46.00 ± 4.17	59.00 ± 4.76	72.00 ± 5.48

Table 4.1: Prediction performance of the AR model reported as Mean±SD. RMSE is in mg/dL, MARD and COD in %, and TD in minutes. Best values are in bold.

Overall, Table 4.1 shows a deterioration in the predictive performance of the model as the PH value increases, as expected from the typical behavior of predictive models. This is related to the fact that BG levels forecasting becomes more challenging as the prediction horizon is extended, due to the increased influence of various factors affecting glucose fluctuations, such as meal intake, insulin administration, physical activity or stress levels. RMSE ranges from 16.73 ± 2.78 mg/dL at 30-min PH to 36.72 ± 9.73 mg/dL at 90-min PH, showing a significant increase for higher PH values. Similarly, MARD varies from 8.11 ± 0.51% for 30 min-PH

to 26.92 ± 14.84% for 90 min-PH. COD shows a marked decrease as the prediction horizon increases, starting from 83.11 ± 7.34% at 30 min and decreasing to 26.92 ± 14.84% at 90 min, suggesting that the model's ability to explain the variance of the data decreases significantly over time. Finally, the Time Delay increases with PH, rising from 21.75 ± 2.45 at 30 min-PH to 72.00 ± 5.48 at 90 min-PH, showing an increasing lag between the actual and predicted glycemic trace for higher prediction horizons, thus indicating that the model's ability to accurately predict glucose levels worsens.

4.2 Long Short-Term-Memory (LSTM)

This section presents the prediction performance of the LSTM model, achieved using the same evaluation metrics and PH values shown for the AR model. Only results obtained from the Multi-Horizon approach are reported, since we have proven that this strategy outperforms the Single-Horizon approach. Table 4.2 illustrates the evaluation metrics achieved by the model, comparing the values obtained for increasing duration of the time series past history (3h, 6h, 12h, and 24h).

Evaluation Metric	Past History	Prediction Horizon				
		30 min	45 min	60 min	75 min	90 min
RMSE [mg/dL]	3h	16.62 ± 2.73	22.94 ± 5.03	28.35 ± 6.97	33.04 ± 9.06	36.66 ± 10.26
	6h	16.46 ± 2.83	22.64 ± 4.82	28.01 ± 6.79	32.52 ± 8.57	36.16 ± 9.95
	12h	16.27 ± 2.60	21.99 ± 4.24	26.79 ± 5.83	30.90 ± 6.87	34.11 ± 8.14
	24h	**15.41 ± 2.50**	**20.38 ± 4.00**	**24.31 ± 5.21**	**27.51 ± 6.21**	**30.28 ± 7.19**
MARD [%]	3h	8 ± 0.49	10.51 ± 0.86	12.94 ± 1.40	15.02 ± 2.00	17.09 ± 2.42
	6h	7.94 ± 0.59	10.64 ± 1.10	13.03 ± 1.68	15.20 ± 2.30	17.17 ± 2.80
	12h	8.12 ± 0.65	10.81 ± 1.25	12.84 ± 1.48	12.84 ± 1.48	16.91 ± 2.68
	24h	**7.65 ± 0.59**	**9.79 ± 0.98**	**11.54 ± 1.36**	**13.06 ± 1.76**	**14.43 ± 2.14**
COD [%]	3h	83.28 ± 7.43	69.78 ± 10.4	55.27 ± 12.33	40.95 ± 12.93	27.98 ± 13.77
	6h	83.79 ± 6.83	70.62 ± 9.88	56.46 ± 11.74	42.71 ± 12.43	30.12 ± 12.78
	12h	83.99 ± 7.07	71.94 ± 9.9	59.22 ± 12.69	46.59 ± 14.37	36.05 ± 14.84
	24h	**85.53 ± 6.52**	**75.6 ± 9.29**	**66.09 ± 11.20**	**57.36 ± 12.31**	**49.12 ± 12.85**
Time Delay [min]	3h	21.50 ± 2.35	32.75 ± 3.43	45.50 ± 4.26	59.00 ± 4.47	70.50 ± 5.10
	6h	21.00 ± 3.08	31.25 ± 3.93	43.75 ± 5.10	54.75 ± 6.38	64.25 ± 7.30
	12h	20.75 ± 3.35	30.25 ± 4.99	40.75 ± 6.13	48.5 ± 6.09	56.25 ± 8.56
	24h	**19.5 ± 3.94**	**26.00 ± 5.98**	**31.25 ± 7.05**	**34.5 ± 7.76**	**35.50 ± 7.24**

Table 4.2: Prediction performance of the LSTM model reported as Mean±SD. Method: Multi-Horizon. Best values per PH are in bold.

In line with the literature [50]-[51]-[52], we observe a decrease in predictive performance for the LSTM model as the prediction horizon increases. We can also state that by extending the past history used for training, the model achieves better performance for all PHs, with the best results observed for a past history of 24 hours. Specifically, RMSE ranges from 16.62 ±

2.73 mg/dL (3h-history) to 15.41 ± 2.5 mg/dL (24h-history) at 30-min PH and from 36.66 ± 10.26 mg/dL (3h-history) to 30.28 ± 7.19 mg/dL (24h-history) at 90-min PH. A similar trend is observed in MARD, which decreases as the past history increases: at 30-min PH, it drops from 8.00 ± 0.49% (3h-history) to 7.65 ± 0.0.59% (24h-history), whereas at 90-min PH, it decreases from 17.09 ± 2.42% (3h-history) to 14.43 ± 2.14% (24h-history). COD shows an improvement with increasing past history, suggesting that a longer time history allows the model to better explain the variance in the data. For instance, at 30 min-PH COD increases from 83.28 ± 7.43% (3h-history) to 85.53 ± 6.52% (24h-history), while at 90-min PH, it rises from 27.98 ± 13.77% (3h-history) to 49.19 ± 12.85% (24h-history). Finally, regarding the Time Delay, we observe a marked reduction as the past history increases: at 30-min PH, it ranges from 21.5 ± 2.35 min (3h-history) to 19.5 ± 3.94 min (24h-history), while at 90-min PH, from 70.5 ± 5.1 min (3h-history) to 35.5 ± 7.2 min (24h-history), further confirming the benefits of considering a longer historical context.

Comparison of AR and LSTM results

The comparison between AR and LSTM models highlights that the LSTM model achieves a better performance for all the evaluation metrics and PHs. First, the RMSE values are lower for the LSTM model. For instance, at 30-min PH, the LSTM model achieves a value of 15.41 ± 2.5 mg/dL (with a past history of 24 hours), while the AR model a value of 16.73 ± 2.78 mg/dL; showing a decrease of 7.89% compared to the AR model. Similarly, the LSTM model demonstrates lower MARD values for all PHs. At 30 min, the best configuration (24h-history) achieves a value of 7.65 ± 0.59%, compared to 8.11 ± 0.51% for the AR model; with a decrease of 5.67%. Furthermore, COD values highlight the superior ability of the LSTM model to explain the variance in the data. While the AR model shows a marked decrease in COD as PH increases, LSTM maintains better results. For instance, at 30 min, the LSTM model shows a value of 85.53 ± 6.52%, compared to a value of 83.11 ± 7.34% for the AR model (increase of 2.91%). Finally, the LSTM model shows a lower Time Delay, indicating that it better captures the dynamics of glucose fluctuations compared to the AR model; at 30 min the LSTM model (19.5 ± 3.94 min) shows a much lower lag than the AR model (21.75 ± 2.45), with an improvement of 10.34%.

The difference in performance between the two models can be further observed in the RMSE boxplots shown in Figure 4.2, highlighting how LSTM obtains lower RMSE values over different prediction horizons than the AR model. Comparing the two boxplots, we can notice that the RMSE distribution for the LSTM model is not only lower on average, but also shows less variability, indicating more stable and accurate forecasts. In contrast, the AR model shows higher RMSE values, with a wider spread. Additionally, as the prediction horizon increases, this difference between the two boxplots becomes even more pronounced, highlighting the superior

performance of the LSTM model over longer forecasting periods.

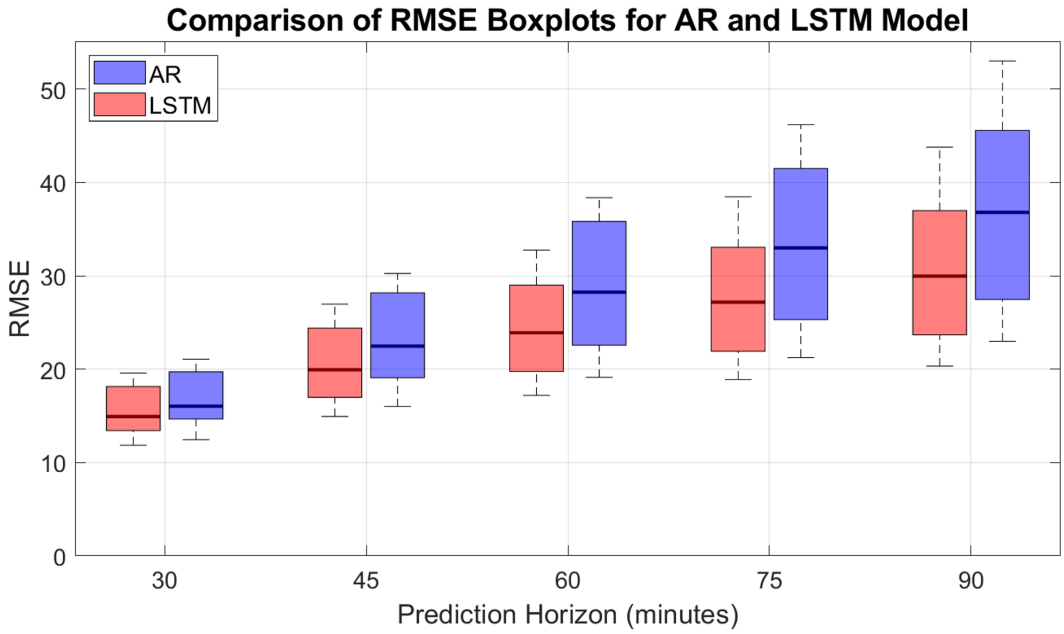


Figure 4.1: RMSE boxplots for all PH values for AR model (in blue) and LSTM model (in red).

4.3 Convolutional Neural Network (CNN)

This section reports the performance metrics obtained for the CNN model using the three image conversion techniques: RP (Subsection 4.3.1), GASF (Subsection 4.3.2), and MTF (Subsection 4.3.3). Again, the evaluation follows the Multi-Horizon prediction strategy by assessing the predictive performance of the model over different lengths of past history (3h, 6h, 12h and 24h). Finally, Section 4.3.4 will present a comparison of the results obtained from all the considered time-series-to-image encoding techniques.

4.3.1 Recurrence Plot

Table 4.3 presents the predictive performance metrics of the CNN model trained using Recurrence Plot images of the CGM.

A key point to highlight is that increasing past history leads to even more significant performance improvement than the LSTM model. For example, the RMSE at 30 min-PH decreases from 30.56 ± 7.08 mg/dL (3h-history) to 21.72 ± 4.38 mg/dL (24h-history), while at 90 min-PH it ranges from 39.24 ± 10.5 mg/dL (3h-history) to 30.51 ± 5.77 mg/dL (24h-history). MARD also shows a marked improvement: at 30 min-PH, this value decreases from 16.81 ± 4.09% (3h-history) to 11.63 ± 3.25% (24h-history), while at 90 min-PH from 19.95 ± 4.25% (3h-history)

Evaluation Metric	Past History	Prediction Horizon				
		30 min	45 min	60 min	75 min	90 min
RMSE [mg/dL]	3h	30.56 ± 7.08	32.57 ± 7.18	34.86 ± 8.14	37.12 ± 9.14	39.24 ± 10.5
	6h	26.72 ± 5.61	29.48 ± 6.00	32.54 ± 6.63	35.16 ± 8.04	37.61 ± 9.06
	12h	24.58 ± 4.02	26.92 ± 4.11	29.43 ± 4.51	31.95 ± 5.13	34.06 ± 5.88
	24h	**21.72 ± 4.38**	**23.94 ± 4.53**	**26.21 ± 4.84**	**28.27 ± 5.26**	**30.51 ± 5.77**
MARD [%]	3h	16.81 ± 4.09	17.51 ± 3.96	18.31 ± 4.06	18.31 ± 4.06	19.95 ± 4.25
	6h	14.11 ± 3.04	15.16 ± 3.12	16.64 ± 3.35	17.66 ± 3.51	18.94 ± 3.82
	12h	12.98 ± 2.63	13.92 ± 2.75	15.02 ± 3.03	16.23 ± 3.38	17.29 ± 3.72
	24h	**11.63 ± 3.25**	**12.45 ± 3.13**	**13.35 ± 3.04**	**14.19 ± 3.02**	**15.26 ± 3.11**
COD [%]	3h	46.22 ± 20.09	39.71 ± 19.37	29.61 ± 17.43	24.8 ± 15.99	17.66 ± 13.61
	6h	57.51 ± 19.46	49.58 ± 19.32	39.37 ± 20.72	31.22 ± 18.96	22.26 ± 19.56
	12h	61.08 ± 23.31	54.49 ± 24.16	46.67 ± 25.61	38.07 ± 27.64	30.7 ± 28.57
	24h	**68.85 ± 24.02**	**63.8 ± 23.03**	**57.82 ± 22.74**	**51.89 ± 22.65**	**44.75 ± 23.52**
Time Delay [min]	3h	13.5 ± 7.45	15.75 ± 7.83	28.25 ± 10.55	37.5 ± 12.09	49 ± 16.59
	6h	12.25 ± 5.25	21.5 ± 7.09	28 ± 11.05	36.75 ± 11.73	44 ± 16.35
	12h	10.5 ± 3.94	14.5 ± 4.56	13.75 ± 7.23	14.75 ± 7.52	17.75 ± 8.66
	24h	**10 ± 3.63**	**10.5 ± 3.59**	**9.5 ± 4.56**	**10 ± 4.59**	**11.25 ± 5.35**

Table 4.3: Prediction performance of the CNN model trained with RP images reported as Mean±SD. Best values per PH are in bold.

to 15.26 ± 3.11% (24h-history). COD follows an upward trend with the increase in past history: at 30 min-PH, it ranges from 46.22 ± 20.09% (3h-history) to 68.85 ± 24.02% (24h-history), while at 90 min-PH from 17.66 ± 13.61% (3h-history) to 44.75 ± 23.52% (24h-history). Finally, Time Delay reveals a significant improvement, with values decreasing progressively for all past-history durations. At 30 min-PH, this value decreases from 13.5 ± 7.45 (3h-history) to 10 ± 3.63 min (24h-history), while at 90 min-PH it decreases from 49 ± 16.59 min (3h-history) to 11.25 ± 5.35 min (24h-history).

4.3.2 Gramian Angular Field

Table 4.4 presents the predictive performance metrics of the CNN model trained using Gramian Angular Field, comparing results obtained with GASF (Table 4.4a) and those obtained with GADF (Table 4.4b). As for the RP technique, also for GAF a significant improvement in the predictive performance is observed as past history increases for all PHs, with the best results obtained for 24-hour.

GASF

Regarding the evaluation metrics for GASF, the RMSE reaches a minimum value with 24h-historty of 21.72 ± 4.38 mg/dL for the 30 min-PH and 30.51 ± 5.77 mg/dL for the 90 min-PH.

For MARD, the minimum value is 11.63 ± 3.25% for 30 min-PH and 15.26 ± 3.11% for 90 min-PH. The COD value reaches the maximum value of 86.05 ± 5.90% for 30 min-PH and 57.45 ± 13.88% for 90 min-PH. Finally, the Time delay shows a minimum values of 14 ± 5.52 min for 30 min-PH and 22 ± 8.94 min for 90 min-PH.

GADF

Concerning the GADF technique, we observe worse performance than GASF in all evaluation metrics, although the general trend of improved performance with increasing past history is observed. Specifically, for RMSE, the values for GADF are higher (18.46 ± 2.82 mg/dL for 30 min-PH and 28.99 ± 5.24 mg/dL for 90 min-PH) compared to the corresponding GASF values, suggesting that the predictions achieved by GADF are less accurate. Similarly, when considering MARD, the GADF technique exhibits higher values (10.16 ± 1.41% for the 30 min-PH and 14.63 ± 2.06% for the 90 min-PH) with respect to those achieved by GASF. Even when looking at the COD value, we observe worse performance of GADF than the GASF technique; i.e., the maximum value achieved by GADF is 78.24 ± 4.55% for 30 min-PH and 60.32 ± 11.24% for 90 min-PH. Finally, differently other metrics, when considering time delay the GADF method shows better values compared to GASF. Specifically, with a 24-hour history, GADF achieves a minimum value of 14.25 ± 5.73 min for 30-min pH and 15.75 ± 5.68 min for 90-min PH. These lower values indicate that GADF generates predictions with a shorter lag than the actual values, allowing for more timely forecasts.

Evaluation Metrics	Past History	Prediction Horizon				
		30 min	45 min	60 min	75 min	90 min
RMSE [mg/dL]	3h	16.06 ± 2.48	22.07 ± 4.27	27.39 ± 5.94	31.81 ± 7.52	35.44 ± 8.78
	6h	16.28 ± 2.42	21.84 ± 4.07	26.85 ± 5.69	31.14 ± 7.11	34.64 ± 8.44
	12h	15.78 ± 2.35	20.49 ± 3.71	24.42 ± 4.37	27.77 ± 4.96	30.65 ± 5.75
	24h	**15.28 ± 2.68**	**19.07 ± 3.45**	**22.21 ± 4.23**	**24.93 ± 4.95**	**27.22 ± 5.66**
MARD [%]	3h	8.03 ± 0.89	10.72 ± 1.32	13.20 ± 1.75	15.49 ± 2.28	17.53 ± 2.79
	6h	8.21 ± 0.88	10.52 ± 1.22	12.84 ± 1.71	15.11 ± 2.29	16.89 ± 2.77
	12h	8.13 ± 0.83	10.10 ± 1.38	11.88 ± 1.89	13.55 ± 2.46	14.94 ± 2.98
	24h	**7.85 ± 1.06**	**9.67 ± 1.41**	**11.16 ± 1.78**	**12.58 ± 2.19**	**13.81 ± 2.57**
COD[%]	3h	84.37 ± 6.73	71.56 ± 10.15	57.30 ± 12.99	43.72 ± 14.18	31.23 ± 14.73
	6h	83.94 ± 6.98	72.08 ± 10.12	59.03 ± 12.22	45.99 ± 13.55	34.35 ± 13.63
	12h	84.87 ± 6.84	75.01 ± 9.94	64.95 ± 10.19	55.23 ± 12.03	46.14 ± 12.29
	24h	**86.05 ± 5.90**	**78.39 ± 8.78**	**71.06 ± 10.83**	**63.93 ± 12.59**	**57.45 ± 13.88**
Time Delay [min]	3h	17.5 ± 3.44)	28.5 ± 5.64	39.75 ± 4.99	51 ± 6.60	61.2 ± 6.25
	6h	17.25 ± 3.79	27.5 ± 5.73	37.5 ± 6.17	47.25 ± 7.15	56.25 ± 7.58
	12h	15.75 ± 3.27	21.25 ± 5.07	26.75 ± 5.91	26.75 ± 6.98	29.75 ± 7.35
	24h	**14 ± 5.52**	**16.75 ± 8.47**	**18.75 ± 9.5**	**19.25 ± 9.21**	**22 ± 8.94**

(a) Prediction performance of the CNN model trained with GASF images.

Evaluation Metrics	Past History	Prediction Horizon				
		30 min	45 min	60 min	75 min	90 min
RMSE	3h	25.06 ± 4.48	27.71 ± 5.36	30.81 ± 6.7	33.81 ± 8.1	36.52 ± 9.35
	6h	21.56 ± 3.47	25.47 ± 4.59	29.34 ± 6.01	32.83 ± 7.4	35.12 ± 8.1
	12h	19.03 ± 2.55	22.42 ± 3.45	25.55 ± 4.29	28.53 ± 5.07	31.26 ± 5.71
	24h	**18.46 ± 2.82**	**21.52 ± 3.35**	**24.31 ± 3.93**	**26.85 ± 4.58**	**28.99 ± 5.24**
MARD [%]	3h	13.95 ± 2.73	14.77 ± 2.72	15.8 ± 2.81	16.91 ± 2.99	18.08 ± 3.21
	6h	11.98 ± 2.09	13.46 ± 2.28	14.99 ± 2.67	16.09 ± 3.07	17.88 ± 3.5
	12h	10.33 ± 1.28	11.56 ± 1.39	12.74 ± 1.67	13.95 ± 1.95	15.24 ± 2.32
	24h	**10.16 ± 1.41**	**11.42 ± 1.46**	**12.58 ± 1.56**	**13.63 ± 1.75**	**14.63 ± 2.06**
COD[%]	3h	68.35 ± 8.12	63.42 ± 9.55	57.12 ± 11.23	50.93 ± 12.87	44.76 ± 13.45
	6h	72.14 ± 7.86	68.09 ± 9.12	62.76 ± 10.15	56.32 ± 11.47	50.12 ± 12.85
	12h	75.87 ± 7.34	71.23 ± 8.96	66.51 ± 9.88	61.15 ± 10.79	55.89 ± 11.72
	24h	**78.24 ± 6.90**	**74.65 ± 8.23**	**70.09 ± 9.35**	**65.47 ± 10.56**	**60.32 ± 11.24**
Time Delay [min]	3h	14.25 ± 5.73	21 ± 8.21	31.25 ± 9.72	36.75 ± 12.27	46.75 ± 12.59
	6h	13 ± 5.71	22.5 ± 7.16	31.5 ± 8.75	42 ± 11.17	51.5 ± 11.6
	12h	12.75 ± 4.44	16.5 ± 6.3	19 ± 6.2	20.25 ± 6.97	21 ± 8.37
	24h	**11.25 ± 4.55**	**13.25 ± 5.45**	**13 ± 4.7**	**14 ± 5.03**	**15.75 ± 5.68**

(b) Prediction performance of the CNN model trained with GADF images.

Table 4.4: Prediction performance of the CNN model trained with GASF and GADF images.

4.3.3 Markov Transition Field

Table 4.5 shows the predictive performance metrics of the CNN model trained with MTF images for the same history values. The RMSE decreases as past history increases, reaching a minimum value of 30.22 ± 5.99 mg/dL (at 30 min-PH) and 35.54 ± 7.09 mg/dL (at 30 min-PH) when considering a time history of 24 hour. Concerning MARD, the lowest value obtained for the 30 min-PH, is 15.64 ± 3.85%, while for the 90 min-PH, the best performance reaches a value of 18.01 ± 3.92%. The maximum COD values achieved is 44.14 ± 30.88% at 30 min-PH, and 26.62 ± 28.1% at 90 min-PH. Finally, regarding Time Delay, the table exhibits a reduction as the past history increases, with the minimum values being 8.5 ± 15.57 min at 30 min-PH and 7.25 ± 6.97 min at 90 min-PH.

Evaluation Metric	Past History	Prediction Horizon				
		30 min	45 min	60 min	75 min	90 min
RMSE [mg/dL]	3h	36.37 ± 8.24	37.83 ± 8.49	39.27 ± 8.91	41.07 ± 9.99	42.23 ± 10.72
	6h	33.32 ± 7.61	35.08 ± 7.81	37.3 ± 8.24	39.53 ± 10.14	41.19 ± 10.31
	12h	31.93 ± 5.81	33.62 ± 5.95	35.52 ± 6.25	37.76 ± 6.6	39.77 ± 6.92
	24h	**30.22 ± 5.99**	**31.63 ± 6.14**	**32.98 ± 6.42**	**34.3 ± 6.77**	**35.54 ± 7.09**
MARD [%]	3h	19.43 ± 4.36	20.10 ± 4.43	20.49 ± 4.29	21.11 ± 4.34	21.73 ± 4.32
	6h	17.11 ± 3.74	18.11 ± 3.98	18.90 ± 3.92	18.9 ± 3.92	21.46 ± 4.52
	12h	16.39 ± 3.5	17.11 ± 3.72	18.24 ± 4.12	19.69 ± 4.57	20.89 ± 4.73
	24h	**15.64 ± 3.85**	**16.18 ± 3.75**	**16.69 ± 3.74**	**17.32 ± 3.72**	**18.01 ± 3.92**
COD [%]	3h	24.99 ± 24.74	19.67 ± 23.35	14.32 ± 21.66	8.04 ± 18.29	3.7 ± 16.79
	6h	37.34 ± 21.06	30.94 ± 21.14	22.14 ± 22.4	15.65 ± 16.02	7.77 ± 19.72
	12h	39.83 ± 22.84	31.93 ± 32.85	24.62 ± 34.08	14.77 ± 38.18	5.75 ± 40.4
	24h	**44.14 ± 30.88**	**39.89 ± 29.39**	**35.34 ± 29.49**	**31.1 ± 28.05**	**26.62 ± 28.1**
Time Delay [min]	3h	14.25 ± 11.15	17.5 ± 11.64	29 ± 10.21	39.5 ± 13.56	55 ± 16.06
	6h	14.25 ± 6.54	25.25 ± 8.5	35.5 ± 10.25	48.75 ± 12.97	54 ± 15.1
	12h	13 ± 9.09	21 ± 7.71	24.75 ± 9.24	25.75 ± 11.27	27.25 ± 11.75
	24h	**8.5 ± 15.57**	**14 ± 6.41**	**11.5 ± 12.58**	**9 ± 7.71**	**7.25 ± 6.97**

Table 4.5: Prediction performance of the CNN model trained with MTF images. Best values per PH are in bold.

4.3.4 Comparison of CNN results

Table 4.6 provides a comparison of evaluation metrics for all image conversion techniques used to train CNN models. The table summarizes the best performance values obtained for each method, i.e., considering a 24-hour past history.

Observing the values reported in the table, it can be concluded that the most performing technique overall is GASF, as evidenced by its significantly better values for the evaluation metrics of RMSE, MARD, and COD. The RMSE obtained with GASF ranges from 15.28 ± 2.68 mg/dL

for a PH of 30 min to 27.22 ± 5.66 mg/dL for a PH of 90 min, resulting the lowest compared to the other techniques. Regarding MARD values, GASF still achieves the best results, with values ranging from 7.85 ± 1.06% at 30 min up to 13.81 ± 2.57% at 90 min. Furthermore, COD reaches the highest values for GASF, decreasing from 86.05 ± 5.90% at 30 min to 57.45 ± 13.88% at 90 min.

Regarding the RMSE, MARD and COD metrics, the least performing technique is MTF. Specifically, the RMSE obtained is the highest among all the techniques, ranging from 30.22 ± 5.99 mg/dL for a PH of 30 min up to 35.54 ± 7.09 mg/dL for a PH of 90 min. Also for MARD, MTF shows the worst values, with a range from 15.64 ± 3.85% at 30 min to 18.01 ± 3.92% at 90 min, resulting significantly higher than the other techniques. Finally, the COD of MTF is the lowest, indicating a lower ability to explain the variability of the data. This value ranges from 44.14 ± 30.88% at 30 min to 26.62 ± 28.1% at 90 min. Concerning the Time Delay, the most performing technique, i.e., the one that achieves the lowest value, depends on the PH value. Specifically, for a PH of 30 min, the minimum value is for GADF, which achieved a Time Delay of 10.25 ± 4.55 min. For a PH of 45 min and 60 min, the lowest values are obtained for RP, with a Time Delay of 10.5 ± 3.59 min and 9.5 ± 4.56 min, respectively. Finally, for longer PHs, i.e., 75 min and 90 min, the technique that achieves the lowest Time Delay is MTF, with values of 9 ± 7.71 min and 7.25 ± 6.97 min.

Evaluation Metric	Encoding Technique	Prediction Horizon				
		30 min	45 min	60 min	75 min	90 min
RMSE [mg/dL]	RP	21.72 ± 4.38	23.94 ± 4.53	26.21 ± 4.84	28.27 ± 5.26	30.51 ± 5.77
	GASF	**15.28 ± 2.68**	**19.07 ± 3.45**	**22.21 ± 4.23**	**24.93 ± 4.95**	**27.22 ± 5.66**
	GADF	18.46 ± 2.82	21.52 ± 3.35	24.31 ± 3.93	26.85 ± 4.58	28.99 ± 5.24
	MTF	30.22 ± 5.99	31.63 ± 6.14	32.98 ± 6.42	34.3 ± 6.77	35.54 ± 7.09
MARD [%]	RP	11.63 ± 3.25	12.45 ± 3.13	13.35 ± 3.04	14.19 ± 3.02	15.26 ± 3.11
	GASF	**7.85 ± 1.06**	**9.67 ± 1.41**	**11.16 ± 1.78**	**12.58 ± 2.19**	**13.81 ± 2.57**
	GADF	10.16 ± 1.41	11.42 ± 1.46	12.58 ± 1.56	13.63 ± 1.75	14.63 ± 2.06
	MTF	15.64 ± 3.85	16.18 ± 3.75	16.69 ± 3.74	17.32 ± 3.72	18.01 ± 3.92
COD [%]	RP	68.85 ± 24.02	63.8 ± 23.03	57.82 ± 22.74	51.89 ± 22.65	44.75 ± 23.52
	GASF	**86.05 ± 5.90**	**78.39 ± 8.78**	**71.06 ± 10.83**	**63.93 ± 12.59**	**57.45 ± 13.88**
	GADF	79.02 ± 9.77	71.67 ± 12.88	64.18 ± 15.7	56.79 ± 18.03	50.42 ± 18.98
	MTF	44.14 ± 30.88	39.89 ± 29.39	35.34 ± 29.49	31.1 ± 28.05	26.62 ± 28.1
Time Delay [min]	RP	10 ± 3.63	**10.5 ± 3.59**	**9.5 ± 4.56**	10 ± 4.59	11.25 ± 5.35
	GASF	14 ± 5.52	16.75 ± 8.47	18.75 ± 9.5	19.25 ± 9.21	22 ± 8.94
	GADF	10.25 ± 4.55	13.25 ± 5.45	13 ± 4.7	14 ± 5.03	15.75 ± 5.68
	MTF	**8.5 ± 15.57**	14 ± 6.41	11.5 ± 12.58	**9 ± 7.71**	**7.25 ± 6.97**

Table 4.6: Comparison of CNN prediction metrics for different encoding techniques (RP, GASF, GADF, MTF) using a 24h past history. Best values per PH are in bold.

4.4 Comparison of LSTM and CNN results

This section discusses the comparison between the performance achieved by the LSTM model and the CNN model trained using the encoding technique shown to perform better, i.e., GASF. Specifically, Table 4.7 shows the evaluation metrics obtained from the two models, considering a 24h-past history.

:

Evaluation Metric	Encoding Technique	Prediction Horizon				
		30 min	45 min	60 min	75 min	90 min
RMSE [mg/dL]	LSTM	15.41 ± 2.5	20.38 ± 4	24.31 ± 5.21	27.51 ± 6.21	30.28 ± 7.19
	CNN (GASF)	**15.28 ± 2.68**	**19.07 ± 3.45**	**22.21 ± 4.23**	**24.93 ± 4.95**	**27.22 ± 5.66**
MARD [%]	LSTM	8 ± 0.49	10.51 ± 0.86	12.94 ± 1.4	15.02 ± 2	17.09 ± 2.42
	CNN (GASF)	**7.85 ± 1.06**	**9.67 ± 1.41**	**11.16 ± 1.78**	**12.58 ± 2.19**	**13.81 ± 2.57**
COD [%]	LSTM	85.53 ± 6.52	75.6 ± 9.29	66.09 ± 11.2	57.36 ± 12.31	49.12 ± 12.85
	CNN (GASF)	**86.05 ± 5.90**	**78.39 ± 8.78**	**71.06 ± 10.83**	**63.93 ± 12.59**	**57.45 ± 13.88**
Time Delay [min]	LSTM	19.5 ± 3.94	26 ± 5.98	31.25 ± 7.05	34.5 ± 7.76	35.5 ± 7.24
	CNN (GASF)	**14 ± 5.52**	**16.75 ± 8.47**	**18.75 ± 9.5**	**19.25 ± 9.21**	**22 ± 8.94**

Table 4.7: Comparison between the performance of the CNN model using the best-performing encoding technique (GASF) and the LSTM model, obtained by considering a 24h-past history. Best values per PH are in bold.

Analyzing all evaluation metrics in Table 4.7, it can be observed that the CNN model, trained using GASF images, exhibits superior performance compared with the LSTM model over all time horizons. Specifically, it can be seen that for short PH values, the two models show very similar prediction performance in terms of RMSE, MARD and COD. Considering a PH of 30 min, LSTM achieved a RMSE value of 15.41 ± 2.5 mg/dL, compared with a value of 15.28 ± 2.68 mg/dL for GASF. Concerning MARD, the values are 8 ± 0.49% for LSTM and 7.85 ± 1.06% for GASF. COD metrics are also comparable: 85.53 ± 6.52% for LSTM and 86.05 ± 5.90% for GASF. In contrast, for longer PHs, the evaluation metrics highlight a more pronounced difference between the performance of the two models. Considering a PH of 90 min the RMSE is 30.28 ± 7.19 mg/dL for LSTM compared with 27.22 ± 5.66 mg/dL for GASF. Similarly, the MARD is 17.09 ± 2.42% for LSTM and 13.81 ± 2.57% for GASF. Furthermore, in terms of COD, the CNN model achieves a value of 57.45 ± 13.88%, while LSTM a value of 49.12 ± 12.85%. The Time Delay of the CNN model using GASF images significantly outperforms that of the LSTM model across all prediction horizons . For instance, at 30 min-PH, the CNN model achieved value of 14 ± 5.52 min, compared to 19.5 ± 3.94 min for LSTM. This difference becomes even more pronounced at a prediction horizon of 90 min, where the CNN model achieves a time delay of 22 ± 8.94 minutes, while LSTM's time delay is 35.5 ± 7.24 minutes. These results indicate that the CNN model with GASF encoding exhibits a much smaller

time shift between the real glucose time series and the predicted values.

This difference in performance is further illustrated in Fig. 4.2, which compares the RMSE boxplots for the LSTM model (in red) and the CNN model (in green) across all considered prediction horizons. From the figure, it can be seen that both boxplots patterns follow a similar rising trend as PH increase. Moreover, while for shorter PH values the RMSE boxplots of the two models appear similar, as the prediction horizon increases the gap between the two boxplots becomes progressively more pronounced. The CNN model not only achieves lower RMSE values compared to the LSTM model for larger PHs, but also exhibits a different distribution spread. Specifically, the RMSE distribution for the CNN model remains more compact and consistent for higher PH values, whereas the LSTM model exhibits a broader spread with larger variability, reflecting greater uncertainty in its predictions as PH increases.

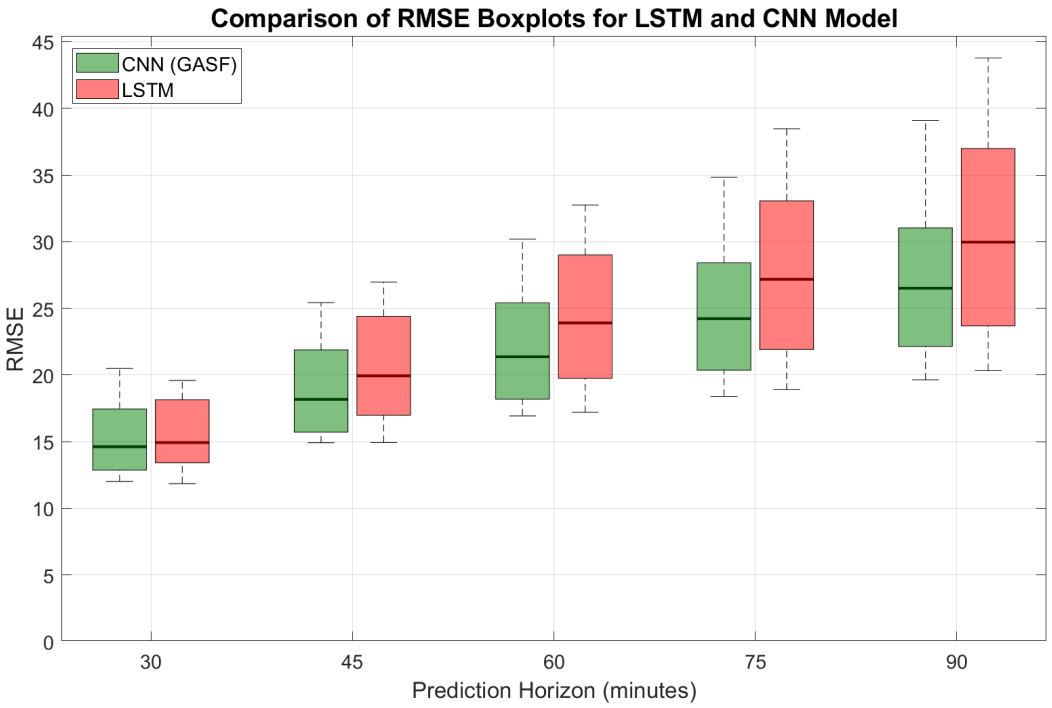


Figure 4.2: RMSE boxplots for all PH values for LSTM model (in blue) and CNN model trained with GASF technique (in red).

Chapter 5

Discussion

This chapter discusses the main findings of this study. Particularly, the results of this thesis offer valuable insights into the blood glucose forecasting capabilities of the CNN model trained with different time series-to-image encoding techniques. In this context, the variation in past history of glycemic time series significantly influenced the visual representation of the generated images and, consequently, the predictive performance of the CNN model, as detailed in Section 5.1. Moreover, the comparison of the CNN model's performance across the three encoding techniques revealed the most effective approach, which will be discussed in Section 5.2. Finally, as will be illustrated in Section 5.3, the superior predictive capability of the CNN model implemented with GASF compared to LSTM was demonstrated. In this chapter, we will not analyze the performance of the Autoregressive model (AR) since it has shown inferior results compared with the LSTM model. Therefore, the LSTM model was chosen as a benchmark for the comparison with the predictions obtained from the CGM-image-based Networks.

5.1 Selection of optimal past history of CGM

In this section, we examine the influence of past history duration on the predictive ability of the CNN model, enabling to determine the optimal length for achieving more accurate BG predictions. To this purpose, different sequence lengths were considered: 36 samples (corresponding to 3 hours), 72 samples (6 hours), 144 samples (12 hours) and 288 samples (24 hours). This analysis allowed us to examine how variability in sequence length affects the predictive performance of the CNN model. Figure 5.1 illustrates the comparison between the original time series and the corresponding images (RP, GAF, and MTF) generated for each sequence length, arranged in increasing order. From Figure 5.1, it can be observed that as the past history increases, the resulting images capture a more comprehensive representation of the glycemic trend. Specifically, shorter sequences tend to produce images with less distinguishable patterns, potentially limiting the model's ability to extract temporal dependencies of the glycemic time series. Conversely, longer sequences generate images that better capture the dynamics underlying BG fluctuations, giving the model a more detailed understanding of the temporal patterns.

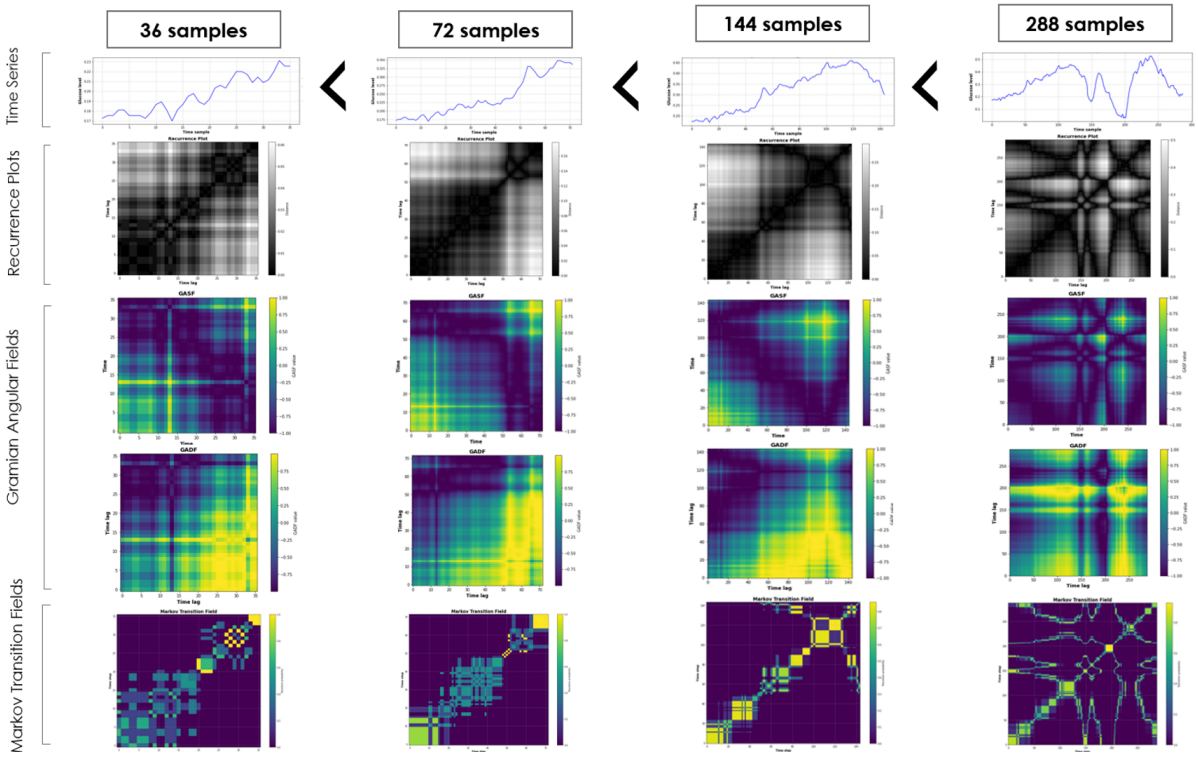


Figure 5.1: Time series and corresponding images (RP, GAF and MTF) generated for each sequence length shown in increasing order, i.e. 36 samples (corresponding to 3h-history) on the left panel, 72 samples (6h-history) on the second panel, 144 samples (12h-history) on the third panel, 288 samples (24-history) on the right panel.

The predictive performance of the three encoding techniques for different sequence lengths, shown in Tables 4.3, 4.4 and 4.5, further support this finding. Regarding the RMSE, a progressive decrease is observed as the past history increases, demonstrating that the model's prediction error reduces with a larger history. For instance, considering a 30-min PH and increasing the input size from 36 samples (3h-history) to 288 samples (24h-history):

- For RP, the RMSE decreases from 30.56 ± 7.08 mg/dL to 21.72 ± 4.38 mg/dL, leading to a median improvement of 11.65%.
- For GASF, the RMSE decreases from 16.06 ± 2.48 mg/dL to 15.28 ± 2.68 mg/dL, with in a median improvement of 3.17%.
- For GADF, the RMSE decreases from 25.06 ± 4.48 mg/dL to 18.46 ± 2.82 mg/dL, resulting in to an improvement of 11.73%.
- For MTF, the RMSE decreases from 36.37 ± 8.24 mg/dL to 30.22 ± 5.99 mg/dL, with a median improvement of 5.36%

For the MARD, a similar progressive decrease is observed as the past history increases, further demonstrating that the model's prediction accuracy improves with the inclusion of more historical data. For instance, considering a 30-min PH and increasing the input size from 36 samples to 288 samples:

- For RP, the MARD decreases from 16.81 ± 4.09% to 11.63 ± 3.25%, leading to a median improvement of 10.43%.
- For GASF, the MARD decreases from 8.03 ± 0.89% to 7.85 ± 1.06%, with a median improvement of 0.87%.
- For GADF, the MARD decreases from 13.95 ± 2.73% to 10.16 ± 1.41%, leading to a median improvement of 13.77%.
- For MTF, the MARD decreases from 19.43 ± 4.36% to 15.64 ± 3.85%, resulting in a median improvement of 4.58%.

Analogously, considering the COD values, a progressive increase is observed as the input history grows, demonstrating an enhancement in the model's ability to explain the variability of the observed CGM values. For instance, considering a 30-min PH and increasing the input size from 36 samples to 288 samples:

- For RP, the COD increases from 46.22 ± 20.09 to 68.85 ± 24.02, leading to a median improvement of 12.72%.
- For GASF, the COD increases from 39.71 ± 19.37 to 63.8 ± 23.03, resulting in a median improvement of 1.11%.
- For GADF, the COD increases from 29.61 ± 17.43 to 57.82 ± 22.74, leading to a median improvement of 5.17%.
- For MTF, the COD increases from 24.8 ± 15.99 to 51.89 ± 22.65, with a median improvement of 10.8%.

Finally, the Time Delay shows a decrease as the input size increases as well, demonstrating the model's ability to make more timely predictions with longer past history data. Considering a 30-min PH and increasing the input size from 36 samples to 288 samples:

- For RP, the time delay decreases from 13.5 ± 7.45 min to 10 ± 3.63 min, leading to a median improvement of 9.26%.
- For GASF, the time delay decreases from 17.5 ± 3.44 min to 14 ± 5.52 min, resulting in a median improvement of 8.70%.

- For GADF, the time delay decreases from 14.25 ± 5.73 min to 11.25 ± 4.55 min, leading to a median improvement of 8.77%.
- For MTF, the time delay decreases from 14.25 ± 11.15 min to 8.5 ± 15.57 min, resulting in a median improvement of 17.07%.

Therefore it can be stated that increasing the length of the past history results in better prediction performance for all encoding techniques, with the greatest median improvement demonstrated for GADF and RP and the smallest for GASF. This suggests that a larger temporal sequence of the glycemic signal improves the ability of the model to recognize temporal patterns and predict future values more effectively. Specifically, the best predictive results were obtained using a past history of 288 samples (24 hours).

5.2 Performance of CGM-image-based Networks

This section describes the comparison between the performance obtained from the three image encoding techniques - RP, GAF and MTF - employed to train the CNN model in BG levels forecasting.

To this end, Figure 5.2 shows a visual representation of the actual CGM values (represented by the dashed blue line) compared with the values predicted by the CNN model trained with the different encoding techniques: RP (in red), GASF (in green), GADF (in purple) and MTF (in yellow). All glycemic profiles shown are obtained using 288-sample input images (24-hour history). First, it can be observed that, considering a 30-min PH (upper panel), all CGM-image-based models are able to capture the fluctuations in glucose levels fairly closely, including peaks (positive and negative), with minimal discrepancy from the actual trend. However, moving to a PH of 90 min (lower panel), the predicted curves follow the target glycemic trend less closely, exhibiting a greater temporal delay. Additionally, predictions struggle to accurately capture hypo- and hyperglycemic events, often leading to an overestimation or underestimation of these peaks. Comparing the predicted curves in Figure 5.2, it can be observed that the GASF technique obtains the most accurate predictions, following the actual glycemic trend more closely and capturing both hypo- and hyperglycemic events more effectively. Conversely, the other three techniques show greater discrepancy in following the actual CGM trend and struggle to accurately capture critical episodes. Among them, MTF emerges as the least effective in accurately reproducing hypo- and hyperglycemic peaks. These findings are particularly evident at the peaks recorded around 09:00 and 14:50, highlighted by the dashed circles in the central panel (PH = 60 min). In these instances, it can be observed that the predictions obtained with the GASF technique are the most accurate in capturing these peaks, although they are not reproduced perfectly, while the other methods are not able to effectively capture them.

Another finding we can draw from Fig. 5.2 is that the predicted traces obtained with the RP and

MTF show a downward offset compared to the target trace, meaning that these two techniques tend to systematically underestimate the actual CGM values. This phenomenon is particularly evident in the time interval between 22:50 and 11:55, where the predicted curves remain consistently lower than the actual values, failing to correctly track glucose dynamics during those periods. Furthermore, it can be observed that these two techniques, RP and MTF, generate noisier glycemic traces, showing more irregular fluctuations and leading to less stable predictions.

These findings are confirmed by the results provided in Section 4.3, which demonstrate that GASF consistently outperforms the other encoding techniques both for short-term (e.g. 30 min-PH) and long-term (e.g. 90 min-PH) predictions. In fact, as shown by the performance metrics in Table 4.6, this method achieves the lowest values of RMSE, with 15.28 ± 2.68 mg/dL at 30-min PH and 27.22 ± 5.66 mg/dL at 90-min PH. Moreover, this technique reaches the minimum MARD value of 7.85 ± 1.06% at 30-min PH and 13.81 ± 2.57% at 90-min PH. Finally, GASF shows the highest COD values, with 86.05 ± 5.90% at 30-min PH and 57.45 ± 13.88% at 90-min PH, further confirming its ability to better capture the glycemic signal compared to the other techniques. GADF follows as the second best method, with good predictive ability but slightly inferior to GASF. The technique that performs the worst overall is MTF, both for lower and higher PHs. As shown in Table 4.6, MTF reaches a RMSE value of 30.22 ± 5.99 mg/dL at 30-min PH and 35.54 ± 7.09 mg/dL at 90-min PH, which are the highest among all techniques. Similarly, for MARD, MTF shows the highest values of 15.64 ± 3.85% at 30-min PH and 18.01 ± 3.92% at 90-min PH. Additionally, this technique achieves the lowest COD values, with 44.14 ± 30.88% at 30-min PH and 26.62 ± 28.1% at 90-min PH.

In summary, from the results obtained from our study, GASF emerges as the best method to obtain accurate predictions, outperforming the other techniques. This suggests that this approach is better able to capture temporal patterns and dynamic changes in the glycemic signal, making it the preferred method for training CNN models in future BG levels forecasting.

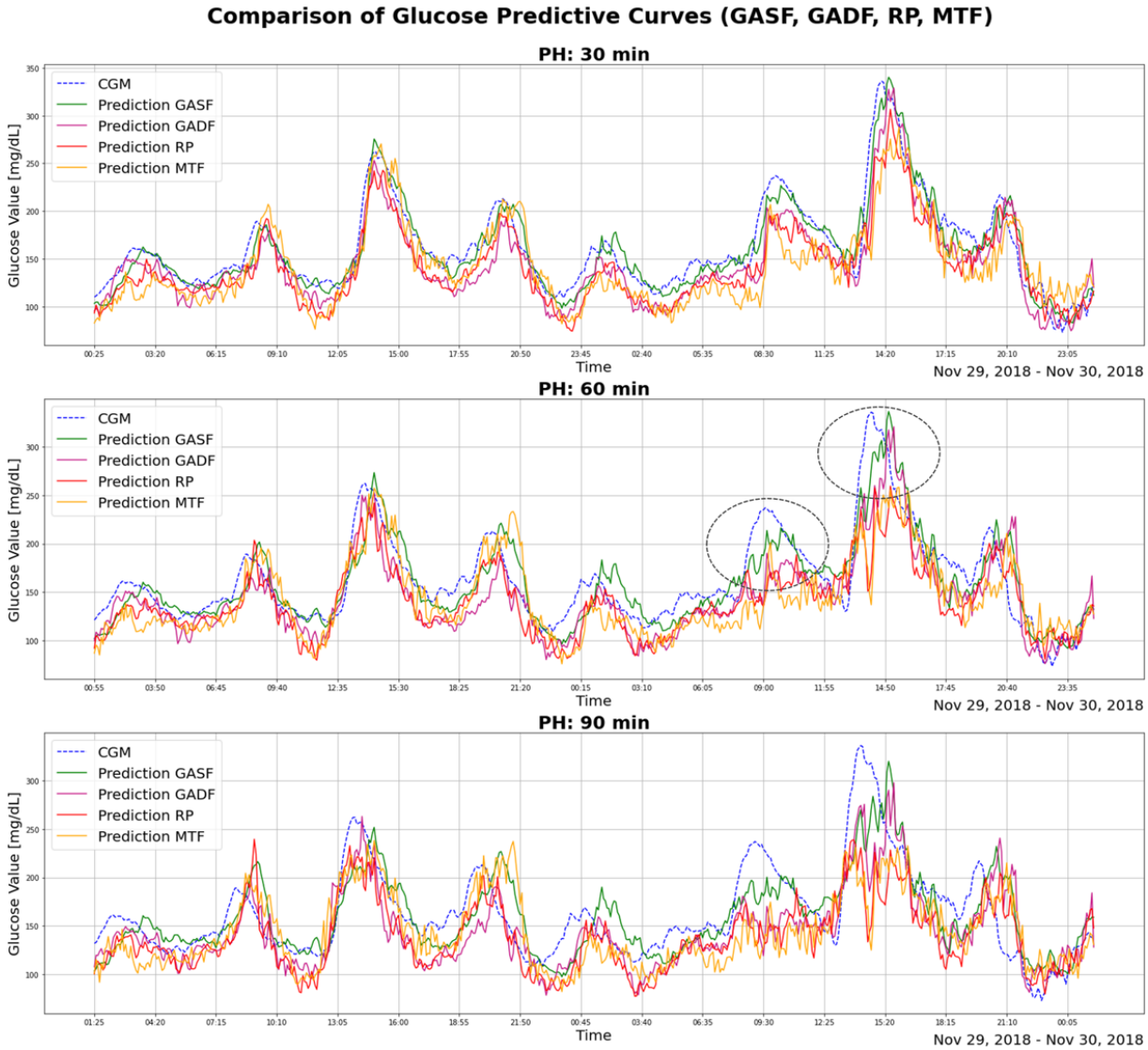


Figure 5.2: Comparison between the actual glucose values (dashed blue line) and those predicted by CNN model with GASF (green line), GADF (purple line), RP (red line), and MTF (yellow line) for test set 12, during a 2-day recording period (from 29 to 30 November). All glycemic profiles are obtained using 288-sample input images (24-hour history).

5.3 Comparison of LSTM and CNN predictions

This section provides a detailed comparison of the performance achieved by LSTM, the benchmark model, and CNN based on GASF which turns out to be the best performing model among the time-series-to-image encoding techniques tested in this work.
For this purpose, Figures 5.3 and 5.4 provide two plots of the comparison between the actual CGM values (represented by the dashed blue line) and the predictions obtained by the LSTM model (in red) and the CNN model trained with the best performing technique, i.e., GASF (in green). All glycemic profiles shown are obtained using 288-sample input images (24-hour history). Observing the predicted glycemic curves, it can be seen that for a short PH (e.g., 30 min in the upper panel), both models track the CGM more closely than for a longer PH (e.g., 90 min in the upper panel). In this latter scenario, the predictions exhibit a reduced ability to accurately capture rapid signal variations and detect hypo- and hyperglycemic events. Comparing the predicted curves of the two models, it can be observed that CNN predicted profiles are closer to the target signal, enabling more accurate and timely predictions than LSTM. In particular, the CNN-based predictions are able to capture hyperglycemic events more effectively. This is particularly evident in the graph related to test set 18 (Fig. 5.3), where two hyperglycemic peaks, recorded around 21:20 (November 29) and 21:10 (November 30), are better captured by the CNN model. On the other hand, at these peaks, the LSTM model struggles to correctly capture the target values, showing a more pronounced time lag, as evidenced by the dashed black circle in the central panel. Similarly, the comparison of the predicted curves for test set 15 (Fig. 5.4) shows that the hyperglycemic episode recorded at 21:00 (November 29) is more accurately predicted by the CNN model, while the LSTM model shows a delay in its forecasts. Regarding hypoglycemic events, both models struggle to capture them accurately; for instance, in the graph shown in Figure 5.4, the hypoglycemic event recorded at 16:00 (Nov 29) is better detected by the LSTM model, while the one occurring at 18:15 (Nov 30) is more closely reproduced by the CNN model (as evidenced by the two rectangles in the central panel).
Moreover, in Figure 5.4 it can be observed that both models tend to overestimate low glucose values and underestimate high glucose values. This is evident for the 90-min PH (lower panel), where the hyperglycemic peaks recorded at 10:10 and 21:50 are underestimated by both models, while the hypoglycemic episodes occurring at 16:00, 02:40 and 18:15 are overestimated. Finally, in Figure 5.3, both predictions appear noisy, particularly in the intervals from 00:55 to 09:00 (Nov 29) and from 00:45 to 12:25 (Nov 30).

These observations are demonstrated by the performance metrics obtained for the two models, as reported in Section 4.4. These results show that the CNN model trained with GASF achieves better performance than LSTM model for all PHs, demonstrating its superior ability to capture both short-term and long-term glycemic trends. For example, comparing the evaluation

metrics obtained at 30 min PH for the subject represented in Figure 5.3 (test set 18), we can observe the following:

- The RMSE achieved by the LSTM model is 13.74 mg/dL and decreases to 12.54 mg/dL for the CNN model, with an improvement of 8.73%.
- The MARD for the LSTM model is 7.40% and increases to 7.00% for the CNN model, showing an decrease of 5.41%.
- The COD for the LSTM model is 84.91% and increases slightly to 85.29% for the CNN model, showing an improvement of 0.44%.
- The Time Delay for the LSTM model is 25 minutes, whereas for the CNN model, it decreases to 20 minutes, with an improvement of 20%.

Moreover, global results provided in Table 4.7 indicate that the CNN model trained with GASF outperforms the LSTM model in all PHs under consideration. In particular, the CNN model obtains lower values of RMSE (15.28 ± 2.68 mg/dL for PH=30 and 27.22 ± 5.66 mg/dL for PH=30) and MARD (7.85 ± 1.06% for PH=30 and 13.81 ± 2.57% for PH=90), highlighting its superior accuracy in predicting future glucose levels. In addition, the CNN model achieves higher COD values (86.05 ± 5.90% for PH=30 and 57.45 ± 13.88% for PH=90), demonstrating its better ability to explain the variance of glucose levels. Finally, the shorter Time Delay obtained by the CNN model trained with GASF (14 ± 5.55 min for PH=30 and 22 ± 8.94 min for PH=90) enables more timely predictions than the LSTM.

In conclusion, the results of our study suggest that transforming CGM time series data into GASF images improves the predictive performance of glycemic forecasting. This time series-to-image encoding method has proven to be a valuable approach for training CNN models in predicting future glycemic levels, improving the predictive accuracy.

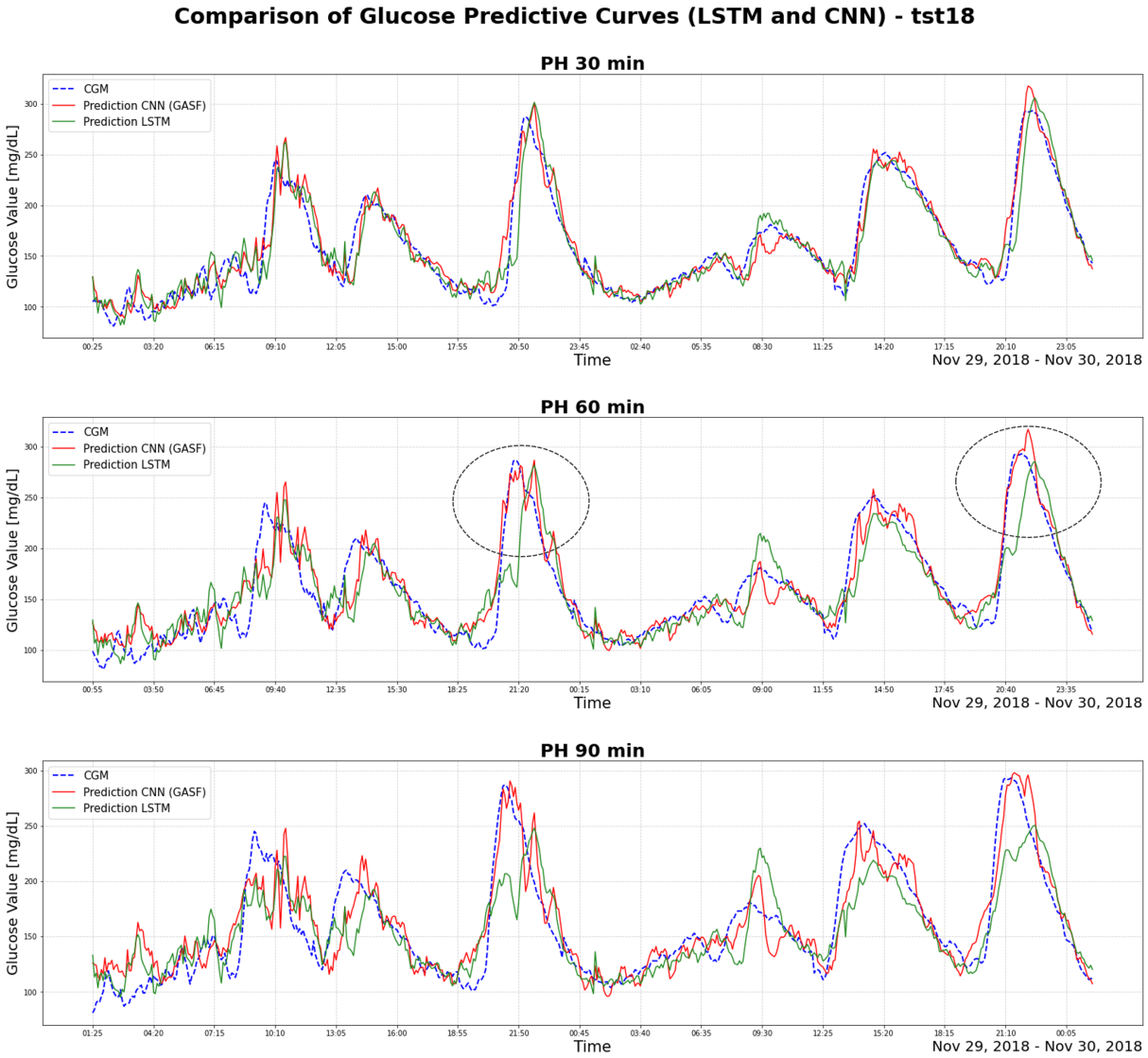


Figure 5.3: Comparison between the actual glucose values (dashed blue line) and those predicted by the LSTM model (red line) and the CNN model with GASF (green line) for test set 18, during a 2-day recording period (from 29 to 30 November). All glycemic profiles are obtained using 288-sample input images (24-hour history).

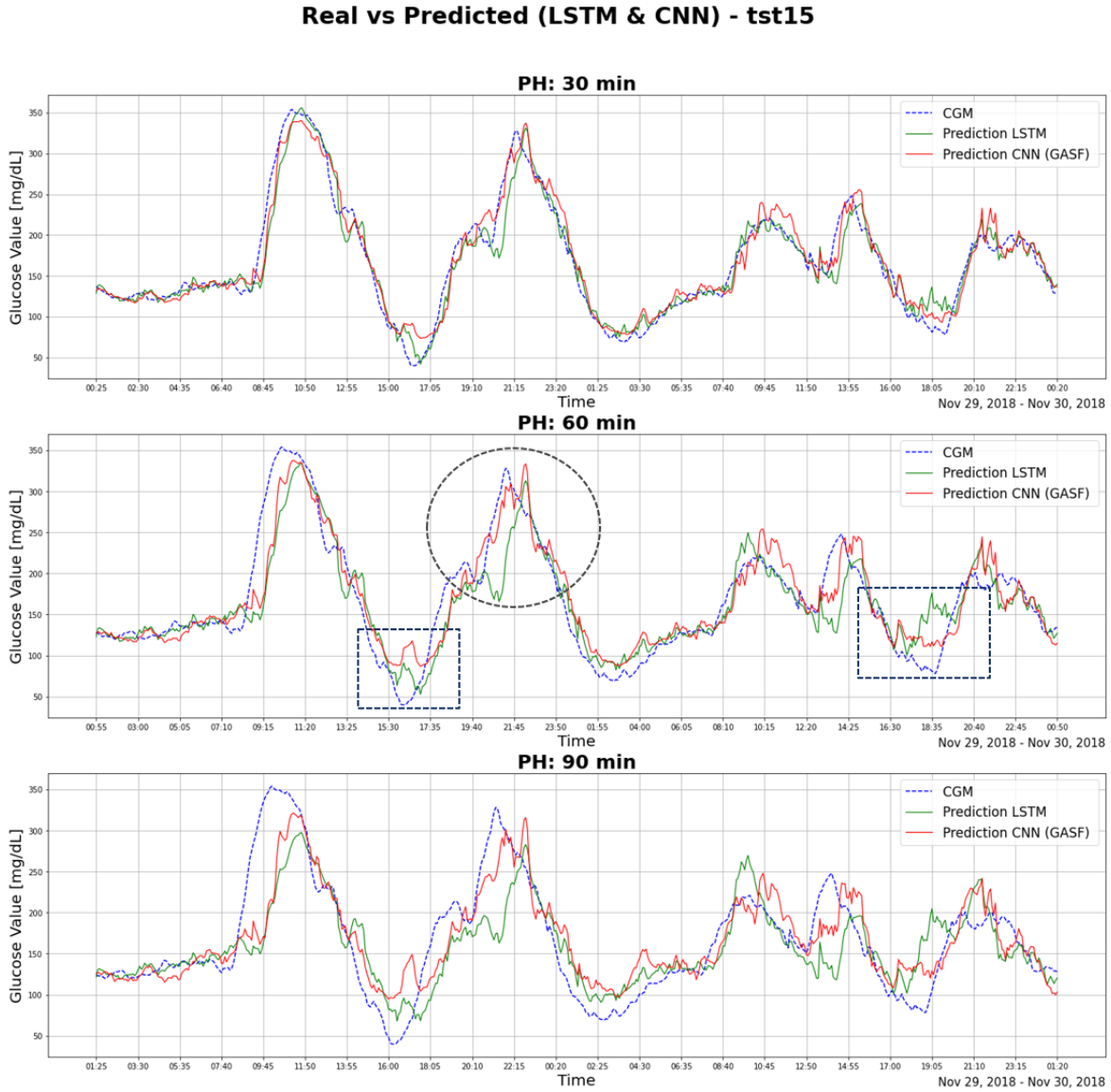


Figure 5.4: Comparison between the actual glucose values (dashed blue line) and those predicted by the LSTM model (red line) and the CNN model with GASF (green line) for test set 15, during a 2-day recording period (from 29 to 30 November). All glycemic profiles are obtained using 288-sample input images (24-hour history).

5.4 Further Analysis on images for different subjects

In this Section, we observe how different temporal patterns in the glycemic data across subjects result in distinct visual structures in the generated matrices, RP, GAF, and MTF. To this purpose, we provide a comparison of the generated matrices for a fixed length of 288 samples (corresponding to 24h past history) from three subjects with different temporal pattern in the glucose time series.

Recurrence Plots

Fig. 5.5 depicts a comparison of the RP images, generated for a 24h-history, of three subjects of the training set, each characterized by a distinct temporal patterns.

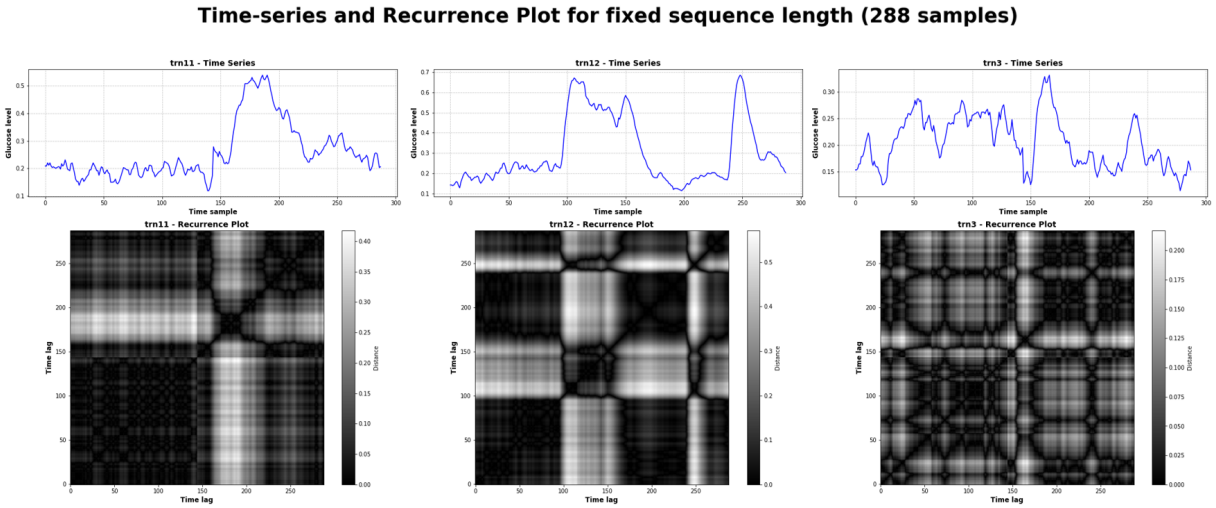


Figure 5.5: Comparison of time-series and RP matrices for three subjects with 288-sample sequences

The time series of training set 11 (left panel) shows a single hyperglycemia peak, ranging from time sample 160 to 220. This appears in the RP matrix as a dark square along the main diagonal, indicating a high similarity in the time series values during this time period. In addition, two light bands (one vertical and one horizontal) emerge, intersecting the dark square, which indicate a decrease in similarity between the peak and values outside that time interval. In the central panel (training set 12), two peaks are evident: the first one longer, extending approximately from time sample 100 to 160, and the second one shorter, localized around time sample 250. In this case, we observe the formation of two vertical and two horizontal light bands, which intersect exactly in the matrix at these two peaks. Finally, the time series of training set 3 (right panel) exhibits a distinct behavior compared to the previous ones, marked by a highly irregular trend with repeated peaks. In the matrix, this irregularity appears as a dense alternation of light and dark regions, creating a "checkerboard" structure that reflects the variability and lack of

regularity in the behavior of the time series.

Gramian Angular Fields

Fig. 5.6 compares the images of three subjects of the dataset, generated using GASF and GADF matrices for a fixed sequences of 288 samples. The time series of the left (training set 32) is characterized by a single peak of hyperglycemia, around sample 190, followed by a drop that reaches its minimum value around sample 260. These features are reflected in the GASF matrix as two yellow spots, corresponding to maximum angular sum values. In the GADF matrix, instead, two yellow stripes intersect at the point related to the positive glycemic peak, indicating regions of increase and decrease in the time series. Regarding the time series of the central panel (training set 12), two main positive peaks and a negative one are observed. These are reflected in the GASF matrix as yellow dots in the main diagonal, and yellow stripes in the GADF matrix. Finally, in the right panel (training set 3) we observe a glycemic signal with an irregular pattern, characterized by many alternating positive and negative peaks. In this case the two matrices appear as a series of yellow dots and areas, forming a kind of 'checkerboard structure'.

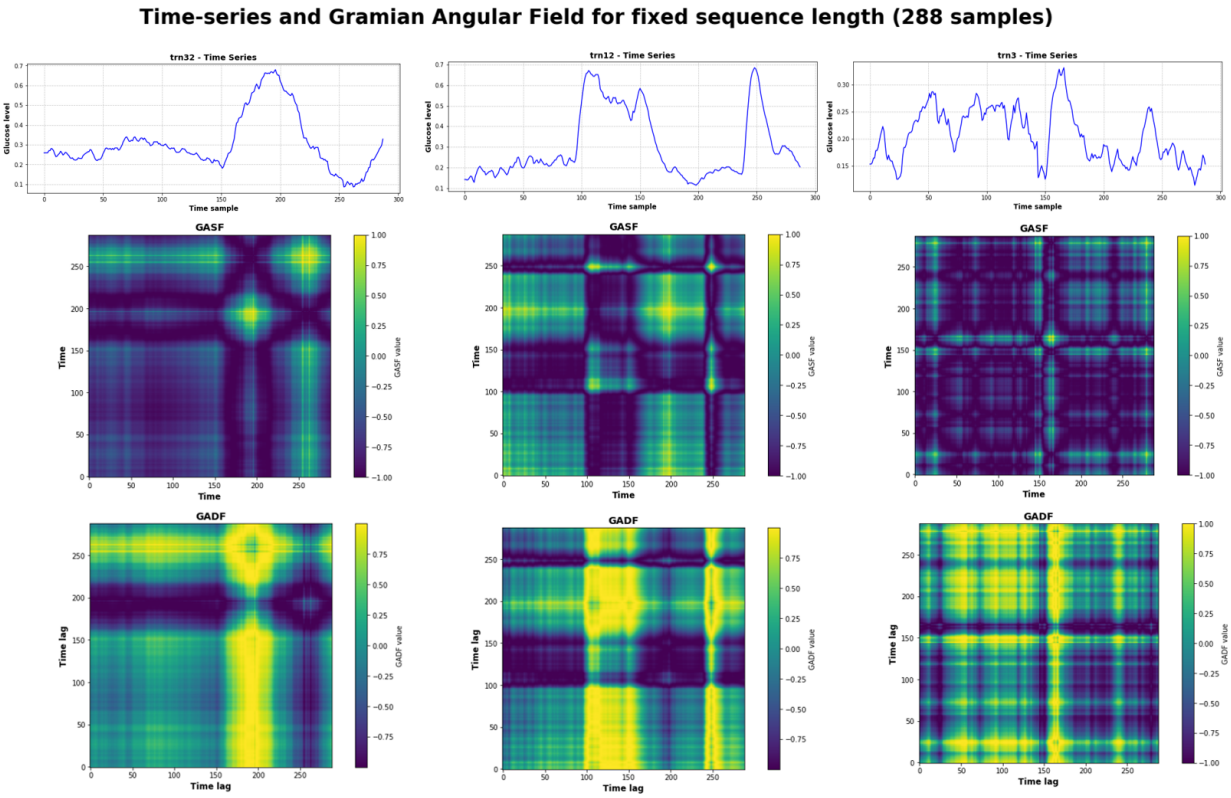


Figure 5.6: Comparison of time-series and GAF matrices for three subjects with 288-sample sequences

Markov Transition Fields

The same analysis and considerations can be conducted for images generated with MTF technique, in this regard subjects of different glycemic trend are compared in Fig. 5.7. Again, it can be observed that the MTF matrices effectively capture variations in transition probabilities, showing unique patterns for different glycemic trends.

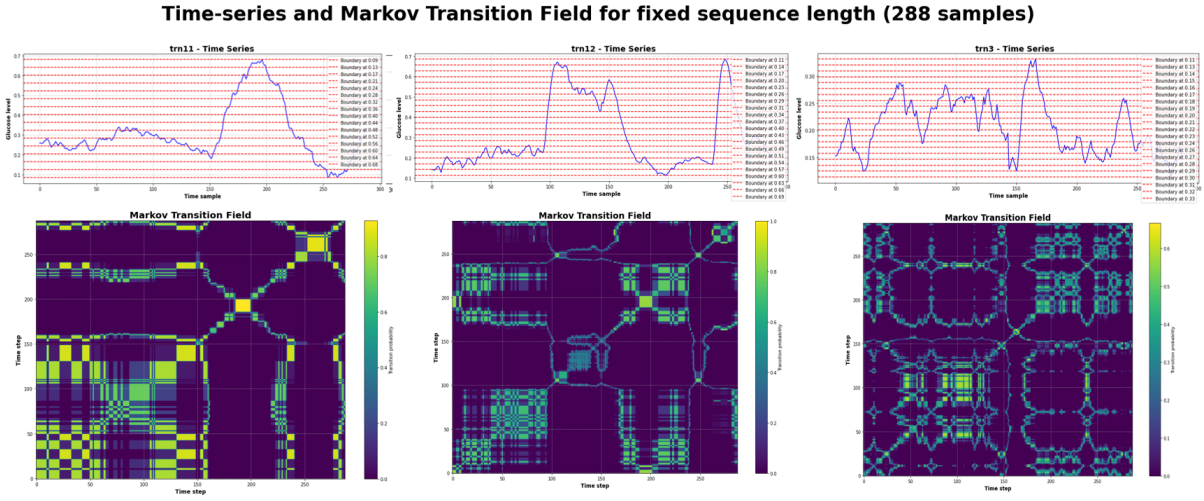


Figure 5.7: Comparison of time-series and MTF matrices for three subjects with 288-sample sequences

In conclusion, from the RP, GAF, and MTF matrices in Figures 5.5, 5.6 and 5.7, we can observe how each T1D subject exhibits a distinct pattern in their BG time series, reflecting individual differences in their glucose dynamics. Our developed CNN model demonstrates a good ability to handle the differences between subjects, as evidenced by the performance metrics obtained for each conversion technique (shown in Tables 4.3, 4.4 and 4.5). However, given the significant variations observed across subjects, developing a customized model for each individual could potentially lead to even better predictive performance, enabling more precise and personalized glucose management for each patient.

Chapter 6

Conclusion and Future Work

Blood glucose (BG) forecasting is a well-established research field that, over the past two decades, has gained significant attention within the T1D diabetes scientific community due to its potential to revolutionize diabetes care [35]. In fact, accurate forecasting of BG levels represents a key element for the development of next-generation tools for the management of T1D therapy, enabling proactive targeted interventions for patients, resulting in reduced risk of acute hypoglycemic or hyperglycemic episodes. Despite the many advances in this area and the large number of published contributions [39]-[50]-[51]-[52], the development of an accurate blood glucose prediction algorithm remains a challenging task because of the complex and highly individualized nature of glucose metabolism.

Inspired by the computer vision area, researchers have begun exploring techniques to transform sequential data into image-like representations, to allow CNNs to fully exploit their potential. Currently, only few contributions deal with this emerging techniques, for instance Barra et al. [55] applied this approach to financial forecasting, Giannetti et al. [56] to weather phenomena classification, and Ko et al.[57] to medical diagnostics. However, this approach has not yet been applied to the development of BG prediction algorithms for T1D.

To fill this gap, we investigated the use of three different techniques - Recurrence Plot, Gramian Angular Field, and Markov Transition Field - to transform the CGM time series in image-like data for developing convolutional neural networks. To understand the impact of the transformation on BG forecasting performance, we conducted a comparative analysis of different input lengths and prediction horizons. Finally, we compare a benchmark Long Short-Term Memory (LSTM) network, which directly processes CGM data, with a deep CNN trained on image-like transformations of CGM sequences.

Chapter 1 presents a brief introduction of T1D complications, management, technologies and an overview of the main state-of-art predictive algorithms.

Chapter 2 introduces the dataset employed in our research, which was generated using the UVA/Padova Type 1 Diabetes Mellitus Simulator (T1DMS). It provides a brief overview of the T1DMS simulation methodology and offers an exploratory analysis of the dataset through the Automated Glucose Data Analysis (AGATA) toolbox.

Chapter 3 provides a detailed explanation of three time-series-image encoding techniques investigated in our study: Recurrence Plot (RP), Gramian Angular Field (GAF) and Markov Transition Field (MTF), describing the methodology used to generate them and analyzing the effect of varying the glycemic past history on their visualization. This chapter also presents an

overview of the three methodologies used for future BG forecasting: an Autoregressive (AR) model and a Long Short-Term Memory (LSTM) network that directly process CGM data, and a Convolutional Neural Network (CNN) trained on the images obtained from our dataset.

Chapter 4 details the results derived from a comparative analysis of different input lengths and prediction horizons, providing important insights into the predictive capabilities of CNN models trained on images, as well as a comparison with the predictive performance of the LSTM model trained on time series data.

Chapter 5 analyzes the main findings obtained from the study, exploring their implications and comparing the performance of the models developed.

Summary of the main findings

The key findings from this study are as follows.

1. The LSTM model consistently outperforms the AR model, achieving significantly better performance across all evaluation metrics (RMSE, MARD, COD, and Time Delay) and for all prediction horizons.

2. Increasing the length of the CGM data as input allows to improve performance for LSTM and CNN based models.

3. The impact of past history length was particularly significant in the predictive performance of deep learning models trained on visual representations. In fact, CNN image-based models exhibited enhanced predictive performance with longer past history, achieving the best results with 24-hour time sequences across all encoding approaches and outperforming models trained on shorter input lengths.

4. Among the image transformation techniques, Gramian Angular Difference Field (GASF) demonstrated the best overall performance, achieving the lowest RMSE and MARD values and the highest COD scores. The predicted CGM profiles demonstrate a better ability to capture hypoglycemic and hyperglycemic episodes than the other techniques.

5. The CNN model trained with the GASF encoding approach outperformed the LSTM model for all PH,particularly in terms of Time Delay.

These results highlight the potential of time-series-to-image encoding techniques in improving the predictive performance of blood glucose levels forecasts through advanced deep learning techniques on a simulated dataset.

Limitations of the study

Despite the promising results highlighted by this study, we acknowledge some limitations that should be considered in future work.First, this is a preliminary investigation that applies an innovative time-series-to-image encoding technique for the development of deep learning models for BG forecasting and a more in-depth analysis is required to test the robustness of the CNN network developed.

Additionally, a key limitation of this work lies in the use of a simulated dataset, which provides an ideal scenario with consistently logged meals and no missing values, outliers, or measurement-related issues. In contrast, real-world clinical datasets often face more challenging situations such as inaccurate meal records, sensor calibration errors, interference from external factors, and delays in sensor response which could affect the performance of predictive models.

Another limitation is that the prediction models employed in this study only considers past glucose history as a parameter for BG forecasting. However, changes in glucose levels are also influenced by other factors, such as carbohydrate intake, insulin dosage, and physical activity. Incorporating these additional factors could potentially lead to even more accurate and robust predictions, as they play a critical role in determining glucose fluctuations.

Finally, we are considering a single model for the entire population, without accounting for the variability between subjects. Taking this variability into consideration could potentially lead to more personalized and accurate BG predictions.

Future work

The results of this study open several perspectives for further research in the field of blood glucose prediction, offering opportunities to enhance forecasting accuracy and expand its applications. Key directions include:

- **Application to real-world datasets:** since our study employed simulated data, validating the proposed approach on real patient CGM data would provide critical insights into its practical applicability.

- **Integration of additional parameters:** incorporating meal intake, physical activity, insulin dosage, and physiological signals could enhance model performance.

- **Exploration of hybrid approaches for time-series-to-image conversion:** combining different encoding techniques, as demonstrated by Wang et al. in [54] who integrated GAF and MTF representations into a single image, could enhance feature extraction and improve predictive performance in BG forecasting.

- **Optimization of deep learning models:** testing alternative deep learning architectures, such as ResNet or hybrid models combining CNN and LSTM as done by Jaloli et al.[52], or Gated Recurrent Units (GRUs) as in [77], could further improve the accuracy and robustness of BG prediction models.
- **Incorporating inter-subject variability:** future work should focus on developing models that account for the variability between subjects, potentially leading to more personalized and accurate blood glucose predictions.
- **Images as model output:** generating GAF images as output from the prediction model and subsequently converting them back into time series could provide a novel approach to enhance the predictive performance.

In conclusion, this study demonstrates the potential of time-series-to-image encoding in BG forecasting and highlights promising directions for further advancements in diabetes management technology.

Bibliography

[1] World Health Organization, *Diabetes*, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes.

[2] National Institute of Diabetes and Digestive and Kidney Diseases, *National institute of diabetes and digestive and kidney diseases*. [Online]. Available: https://www.niddk.nih.gov/.

[3] MSD Manual. "Diabetes mellitus in pregnancy." (2025), [Online]. Available: https://www.msdmanuals.com/professional/gynecology-and-obstetrics/pregnancy-complicated-by-disease/diabetes-mellitus-in-pregnancy.

[4] M. Fowler *et al.*, "Hyperglycemic crisis in adults: Pathophysiology, presentation, pitfalls, and prevention," *Clinical Diabetes*, vol. 27, no. 1, pp. 19–23, Jan. 2009. doi: 10.2337/diaclin.27.1.19.

[5] National Diabetes Data Group, "Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance," *Diabetes*, vol. 28, pp. 1039–1057, 1979. doi: 10.2337/diab.28.12.1039.

[6] K. Alberti *et al.*, "Definition, diagnosis and classification of diabetes mellitus and its complications. part 1: Diagnosis and classification of diabetes mellitus provisional report of a who consultation," *Diabetic Medicine*, vol. 15, no. 7, pp. 539–53, 1998. doi: 10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S.

[7] P. Choudhary *et al.*, "Hypoglycaemia in type 1 diabetes: Technological treatments, their limitations and the place of psychology," *Diabetologia*, vol. 61, no. 4, pp. 761–769, 2018. doi: 10.1007/s00125-018-4566-6.

[8] G. Cappon, "Open-loop insulin dosing personalization in type 1 diabetes using continuous glucose monitoring data and patient characteristics," Ph.D. dissertation, Università degli Studi di Padova, 2020. [Online]. Available: https://www.research.unipd.it/bitstream/11577/3425798/1/cappon_giacomo_thesis.pdf.

[9] J. Kirk *et al.*, "Self-monitoring of blood glucose: Practical aspects," *Journal of Diabetes Science and Technology*, vol. 4, no. 2, pp. 435–439, Mar. 1, 2010. doi: 10.1177/193229681000400225.

[10] E. Benjamin *et al.*, "Self-monitoring of blood glucose: The basics," *Clinical Diabetes*, vol. 20, no. 1, pp. 45–47, Jan. 2002. doi: 10.2337/diaclin.20.1.45.

[11] E. Boland *et al.*, "Limitations of conventional methods of self-monitoring of blood glucose: Lessons learned from 3 days of continuous glucose sensing in pediatric patients with type 1 diabetes," *Diabetes Care*, vol. 24, no. 11, pp. 1858–1862, Nov. 2001. doi: 10.2337/diacare.24.11.1858.

[12] M. Erbach *et al.*, "Interferences and limitations in blood glucose self-testing: An overview of the current knowledge," *Journal of Diabetes Science and Technology*, vol. 10, no. 5, pp. 1161–1168, 2016. doi: 10.1177/1932296816641433.

[13] Medtronic Diabetes, *Monitoraggio continuo del glucosio (cgm)*, 2025. [Online]. Available: https://www.medtronic-diabetes.com/it-IT/conoscere-il-diabete/monitoraggio-continuo-del-glucosio.

[14] National Institute of Diabetes and Digestive and Kidney Diseases, *Continuous glucose monitoring*, 2023. [Online]. Available: https://www.niddk.nih.gov/health-information/diabetes/overview/managing-diabetes/continuous-glucose-monitoring.

[15] G. Cappon *et al.*, "Continuous glucose monitoring sensors for diabetes management: A review of technologies and applications," *Diabetes Metab J*, vol. 43, no. 4, pp. 383–397, 2019. doi: 10.4093/dmj.2019.0121.

[16] Medtronic Diabetes, *Minimed™ 780g insulin pump system*, 2024. [Online]. Available: https://www.medtronic-diabetes.com/it-IT/terapia-con-microinfusore/sistema-integrato.

[17] V. Valla *et al.*, "Continuous subcutaneous insulin infusion (csii) pumps," Advances in Experimental Medicine and Biology, vol. 771, pp. 414–419, 2012. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4614-5441-0_29.

[18] Policlinico Campus Biomedico, *Somministrazione di insulina tramite microinfusore*, 2024. [Online]. Available: https://www.policlinicocampusbiomedico.it/cure/somministrazione-di-insulina-tramite-microinfusore.

[19] Omnipod, *Pod uni insulin pump*, 2024. [Online]. Available: https://www.omnipod.com/en-gb/diabetes-hub/pod-uni/insulin-pump.

[20] S. Sy *et al.*, "Can smart pens help improve diabetes management?" *J Diabetes Sci Technol*, vol. 16, no. 3, pp. 628–634, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9294588/.

[21] S. Kalra *et al.*, "Optimising insulin injection techniques to improve diabetes outcomes," *Diabetes Ther.*, vol. 14, no. 11, pp. 1785–1799, 2023. doi: 10.1007/s13300-023-01460-y..

[22] A. Slomski, "Artificial pancreas improves glycemic control," *JAMA*, vol. 320, no. 20, pp. 2068–2068, 2018. doi: 10.1001/jama.2018.18248.

[23] M. Nwokolo *et al.*, "The artificial pancreas and type 1 diabetes," *The Journal of Clinical Endocrinology & Metabolism*, vol. 108, no. 7, pp. 1614–1623, 2023. doi: 10.1210/clinem/dgad068.PMID:36734145.

[24] B. W. Bequette *et al.*, "Challenges and recent progress in the development of a closed-loop artificial pancreas," *Diabetes Technology & Therapeutics*, vol. 15, no. 8, pp. 682–694, 2013. doi: 10.1016/j.arcontrol.2012.09.007.

[25] R. T. Sutton *et al.*, "An overview of clinical decision support systems: Benefits, risks, and strategies for success," *NPJ Digital Medicine*, vol. 3, p. 17, 2020. doi: 10.1038/s41746-020-0221-y.

[26] I. Contreras *et al.*, "Artificial intelligence for diabetes management and decision support: Literature review," *Journal of Medical Internet Research*, vol. 20, no. 5, e10775, 2018. doi: 10.2196/10775.

[27] G. Noaro *et al.*, "Machine-learning based model to improve insulin bolus calculation in type 1 diabetes therapy," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 247–255, 2020. doi: 10.1109/TBME.2020.2962969.

[28] T. Zhu *et al.*, "An insulin bolus advisor for type 1 diabetes using deep reinforcement learning," *Sensors*, vol. 20, no. 18, p. 5058, 2020. doi: 10.3390/s20185058.

[29] M. Anthimopoulos *et al.*, "Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones," *Journal of Diabetes Science and Technology*, vol. 9, no. 3, pp. 507–515, 2015. doi: 10.1177/1932296814565242.

[30] S. Faccioli *et al.*, "Super–twisting-based meal detector for type 1 diabetes management: Improvement and assessment in a real-life scenario," *Computer Methods and Programs in Biomedicine*, vol. 219, p. 106 736, 2022. doi: 10.1016/j.cmpb.2022.106736.

[31] K. Turksoy *et al.*, "Classification of physical activity: Information to artificial pancreas control systems in real time," *Journal of Diabetes Science and Technology*, vol. 9, no. 6, pp. 1200–1207, 2015. doi: 10.1177/1932296815604246.

[32] K. Turksoy and j. I. v. n. p. y. d. T. others title = Real-time model-based fault detection of continuous glucose sensor measurements,

[33] M. H. Jensen *et al.*, "Professional continuous glucose monitoring in subjects with type 1 diabetes: Retrospective hypoglycemia detection," *Journal of Diabetes Science and Technology*, vol. 7, no. 1, pp. 135–143, 2013. doi: 10.1177/193229681300700115.

[34] P. P. San *et al.*, "Deep learning framework for detection of hypoglycemic episodes in children with type 1 diabetes," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2016, pp. 3503–3506. doi: 10.1109/EMBC.2016.7591565.

[35] Z Guan *et al.*, "Artificial intelligence in diabetes management: Advancements, opportunities, and challenges," *Cell Reports Medicine*, vol. 4, no. 10, p. 101 213, 2023. doi: 10.1016/j.xcrm.2023.101213.

[36] Y. Shen *et al.*, "Personalized blood glucose forecasting from limited cgm data using incrementally retrained lstm," *IEEE Transactions on Biomedical Engineering*, pp. 1–12, 2024. doi: 10.1109/TBME.2024.3494732.

[37] M. Gadaleta *et al.*, "Prediction of adverse glycemic events from continuous glucose monitoring signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 650–659, 2019. doi: 10.1109/JBHI.2018.2823763.

[38] G. Sparacino *et al.*, "Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 931–937, 2007. doi: 10.1109/TBME.2006.889774.

[39] G. Sparacino *et al.*, "Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 931–937, 2007. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/17518291/.

[40] J. Reifman *et al.*, "Predictive monitoring for improved management of glucose levels," *Journal of Diabetes Science and Technology*, vol. 1, no. 4, pp. 478–486, 2007. [Online]. Available: https://europepmc.org/article/MED/19885110.

[41] A. Gani *et al.*, "Universal glucose models for predicting subcutaneous glucose concentration in humans," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 157–165, 2009. doi: 10.1109/TITB.2009.2034141.

[42] D. N. Katsarou *et al.*, "Short term glucose prediction in patients with type 1 diabetes mellitus," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 329–332. doi: 10.1109/EMBC48229.2022.9870889.

[43] M. Otoom *et al.*, "Real-time statistical modeling of blood sugar," *J. Med. Syst.*, vol. 39, no. 10, p. 123, 2015. doi: 10.1007/s10916-015-0301-8.

[44] D. Finan *et al.*, "Experimental evaluation of a recursive model identification technique for type 1 diabetes," *Journal of Diabetes Science and Technology*, vol. 3, no. 5, pp. 1192–1202, 2009. doi: 10.1177/193229680900300526.

[45] M. Eren-Oruklu *et al.*, "Adaptive system identification for estimating future glucose concentrations and hypoglycemia alarms," *Automatica (Oxf)*, vol. 48, no. 8, pp. 1892–1897, 2012. doi: 10.1016/j.automatica.2012.05.076. eprint: 2012Jun22.

[46] M. G. Pedersen, *Machine learning for bioengineering*, INP9087820, Lectures: Random Forest and Support Vector Machines, 2023-24.

[47] E. I. Georga *et al.*, "A predictive model of subcutaneous glucose concentration in type 1 diabetes based on random forests," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 2889–2892. doi: 10.1109/EMBC.2012.6346567.

[48] E. I. Georga *et al.*, "Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 71–81, 2013. doi: 10.1109/TITB.2012.2219876.

[49] I. H. Sarker *et al.*, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, p. 420, 2021. doi: 10.1007/s42979-021-00815-1.

[50] T. Zhu, *et al.*, "A deep learning algorithm for personalized blood glucose prediction," pp. 64–78, 2018. [Online]. Available: https://ceur-ws.org/Vol-2148/paper12.pdf.

[51] K. Li *et al.*, "Glunet: A deep learning framework for accurate glucose forecasting," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 414–423, 2020. doi: 10.1109/JBHI.2019.2931842.

[52] M. Jaloli *et al.*, "Long-term prediction of blood glucose levels in type 1 diabetes using a cnn-lstm-based deep neural network," *Journal of Diabetes Science and Technology*, vol. 17, no. 6, pp. 1590–1601, 2023. doi: 10.1177/19322968221092785.

[53] Z. Zhen *et al.*, "From pixels to predictions: Spectrogram and vision transformer for better time series forecasting," *arXiv preprint arXiv:2403.11047*, 2024, Published at ACM ICAIF 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2403.11047.

[54] Z. Wang *et al.*, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," Jan. 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:16409971.

[55] S. Barra, S. M. Carta, A. Corriga, A. S. Podda, and D. R. Recupero, "Deep learning and time series-to-image encoding for financial forecasting," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 3, pp. 683–692, 2020. doi: 10.1109/JAS.2020.1003132.

[56] C. Giannetti, "A novel time series-to-image encoding approach for weather phenomena classification," Jun. 2024. doi: https://doi.org/10.48550/arXiv.2406.05096.

[57] D.-W. Ko *et al.*, "Schizophrenia diagnosis based on eeg through time-series image conversion and deep learning," *Electronics*, vol. 11, no. 14, p. 2265, 2022. doi: 10.3390/electronics11142265.

[58] C. D. Man *et al.*, "The uva/padova type 1 diabetes simulator: New features," *Journal of Diabetes Science and Technology*, vol. 8, no. 1, pp. 26–34, 2014, Epub 2014 Jan 1. doi: 10.1177/1932296813514502.

[59] T. Team, *Uva/padova type 1 diabetes simulator*, Accessed: 2024-12-03, n.d. [Online]. Available: https://tegvirginia.com/t1dms/.

[60] G. Cappon, *Agata - automated glucose data analysis toolbox*, https://github.com/gcappon/agata, 2025.

[61] G. Cappon *et al.*, "Agata: A toolbox for automated glucose data analysis," *Journal of Diabetes Science and Technology*, vol. 18, no. 5, pp. 1109–1121, 2024. doi: 10.1177/19322968221147570.

[62] *Transform time series into images — pyts 0.12.0 documentation*, Accessed: 2024-11-07. [Online]. Available: https://pyts.readthedocs.io/en/stable/modules/image.html.

[63] Z. Wang *et al.*, "Partially encoding temporal correlations to classify temporal data using convolutional neural networks," *arXiv*, vol. arXiv:1509.07481, 2015. doi: 10.48550/arXiv.1509.07481.

[64] IBM, *Autoregressive model*, 2025. [Online]. Available: https://www.ibm.com/it-it/topics/autoregressive-model.

[65] P. A. Facchinetti, *Elaborazione di segnali biologici*, INL1001851, PDF 04, modelli ARIMA, 2023.

[66] S. Pang, J. Liu, *et al.*, "A photovoltaic power predicting model using the differential evolution algorithm and multi-task learning," *Frontiers in Materials*, vol. 9, p. 938 167, 2022. doi: 10.3389/fmats.2022.938167.

[67] S. Hochreiter *et al.*, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. doi: 10.1162/neco.1997.9.8.1735.

[68] A. Aliberti *et al.*, "A multi-patient data-driven approach to blood glucose prediction," *IEEE Access*, vol. 7, pp. 69 311–69 325, 2019. doi: 10.1109/ACCESS.2019.2919184.

[69] M. F. Rabby *et al.*, "Stacked lstm based deep recurrent neural network with kalman smoothing for blood glucose prediction," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 101, 2021. doi: 10.1186/s12911-021-01462-5.

[70] T. E. Idrissi *et al.*, "Strategies of multi-step-ahead forecasting for blood glucose level using lstm neural networks: A comparative study," in *International Conference on Health Informatics*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:215597322.

[71] F. Prendin *et al.*, "Forecasting of glucose levels and hypoglycemic events: Head-to-head comparison of linear and nonlinear data-driven algorithms based on continuous glucose monitoring data only," *Sensors*, vol. 21, no. 5, 2021. doi: 10.3390/s21051647.

[72] C. Mosquera-Lopez *et al.*, "Incorporating glucose variability into glucose forecasting accuracy assessment using the new glucose variability impact index and the prediction consistency index: An lstm case example," *Journal of Diabetes Science and Technology*, vol. 16, no. 1, pp. 7–18, 2022, issn: 1932-2968. doi: 10.1177/19322968211042621.

[73] D. Nasir *et al.*, "Predicting blood glucose levels in type 1 diabetes using lstm," in *Modern Artificial Intelligence and Data Science: Tools, Techniques and Systems*, Springer, 2023, pp. 121–135.

[74] S. B. Taieb *et al.*, "A bias and variance analysis for multistep-ahead time series forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 1, pp. 62–76, 2015. doi: 10.1109/TNNLS.2015.2411629.

[75] GeeksforGeeks, *Introduction to convolution neural network*, 2023. [Online]. Available: https://www.geeksforgeeks.org/introduction-convolution-neural-network/#cnn-architecture.

[76] K. O'Shea *et al.*, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015. doi: https://doi.org/10.48550/arXiv.1511.08458.

[77] R. I. Alkanhel *et al.*, "Hybrid cnn-gru model for real-time blood glucose forecasting: Enhancing iot-based diabetes management with ai," *Sensors*, vol. 24, no. 23, 2024, issn: 1424-8220. doi: 10.3390/s24237670. [Online]. Available: https://www.mdpi.com/1424-8220/24/23/7670.