الجمهورية الجزائرية الديمقراطية الشعبية

The People's Democratic Republic of Algeria

وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research

جامعة محمد بوضياف بالمسيلة

University Mohamed Boudiaf of M'sila

1985

جامعة محمد بوضياف - المسيلة

Université Mohamed Boudiaf - M'sila

كلية الرياضيات والإعلام الآلي

Faculty of Mathematics and Informatics

قسم الإعلام الآلي

Department of Computer Science

Domain: Mathematics and Computer Science

Thesis Presented to Fulfill the Partial Requirement
for **Master's Degree** in Computer Science

Specialty: Networks and Information and Communication Technologies

Prepared By: Houichi Fadoua

Supervised By:

Lamri Sayad

ENTITLED

DIABETIC BLOOD GLUCOSE PREDICTION USING MACHINE LEARNING

Jury Members

ALLAOUA HEMMAK	President
LAMRI SAYAD	Supervisor
MAROUANE KIHAL	Examiner

Academic Year 2024/2025

Dedication

اهدي ثمرة جهدي الى من اوصاني الله بهما برا اطال الله عمرهما والبسهم لباس الصحة والعافية
الى من كانو خير سند اخوتي الاعزاء ادامكم الله ضلعا ثابتا لي
الى من علمني حرفا طيلة مساري الدراسي ولم يبخل بالعطاء اساتذتي الافاضل

الى نفسي التي راهنت على النجاح

ولكل من اتسع لهم صدري وضاقت هذه الورقة عن ذكرهم
اهديكم عملي المتواضع عرفانا بالجميل وتقديرا لمجهودكم

Acknowledgement

الحمدلله الذي انشا وبرى وخلق الماء والثرى الرحمان على العرش استوى والصلاة على الحبيب المصطفى اما بعد

باسم الله الرحمان الرحيم

" قل اعملو فسيرى الله عملكم ورسوله و المؤمنين"

الاساتذة اعضاء لجنة المناقشة
الاستاذ المشرف الفاضل
الحضور الكرام

السلام عليكم ورحمة الله وبركاته
احمد الله تعالى على توفيقه وفضله الذي لولاه ما وصلت الى هذه المرحلة من مشواري العلمي
يسعدني ويشرفني ان اناقش معكم اليوم مذكرتي لنيل شهادة الماستر راجيتا ان تنال رضاكم وتقديركم

اتوجه بالشكرالجزيل الى اعضاء اللجنة الموقرين
ولاستاذي المشرف على توجيهه
ولعائلتي الكريمة التي كانت سندي في كل خطوة
كما اشكركل من حضر وشاركني هذه اللحضة المهمة كل باسمه ومقامه
بارك الله فيكم ومرحبا بكم

انتهت الرحلة بعون الله وبفضله

" واخر دعواهم ان الحمدلله رب العالمين"
الحمدلله عند البدئ وعند الختام

Table of Contents

List of Figures

List of Tables

List of Abbreviations

T1D Type 1 diabetes

BG Blood Glucose

CGM Continuous Glucose Monitoring

GI Glycemic Index

PH Prediction Horizon

ML Machine Learning

ANN Artificial Neural Network

FFNN Feed Forward Neural Network

RNN Recurrent Neural Network

LSTM Long Short-Term Memory

SVM Support Vector Machine

RMSE Root Mean Square Error

CEGA Clark Error Grid Analysis

GRU Gated Recurrent Unit

General Introduction

Diabetes is a chronic health condition that affects millions of people worldwide. It results from the body's inability to effectively regulate blood glucose levels, leading to deterioration in health if left untreated. Predicting blood glucose levels is crucial for diabetics as it helps them manage their condition and avoid sudden spikes or drops in blood sugar levels. However, traditional blood glucose measurement methods are often inconvenient and require frequent manual intervention.

With advances in artificial intelligence (AI) and deep learning (DL), it has become possible to develop models that can accurately predict blood glucose levels based on historical patient data. These models enable patients to make proactive decisions about their diet and medication based on accurate predictions, improving their quality of life and reducing health risks.

Researchers have explored innovative approaches to accurately predict these levels, as the current preferred treatment relies primarily on self-management, which involves actively tracking blood glucose levels, managing physical activity, diet, and insulin dosage. Recent advances in diabetes technology and self-management applications have made it easier for patients to access relevant data. This report provides a comprehensive analysis of a project designed to predict blood glucose levels using machine learning techniques. During the research phase, we reviewed numerous scientific papers related to this field, as well as several open-source projects aimed at predicting glucose levels. Among these projects, the one developed by Ladislav Floriš and available through the GitHub repository "Blood-Glucose-Prediction" caught our attention due to its satisfactory results. While this foundation provided a solid starting point, we implemented several improvements to enhance its functionality and user experience. These improvements include an improved interactive interface, additional data analysis features, and performance improvements to increase model accuracy.

Motivation

The motivation behind this project stems from the growing prevalence of diabetes globally. According to the World Health Organization (WHO), diabetes was the ninth leading cause of death in 2019, with approximately 1.5 million deaths directly attributed to the disease. Controlling blood glucose levels is essential to improving patients' quality of life and reducing the risk of complications. Machine learning offers a promising solution to this challenge by enabling the development of predictive models that analyze historical data to predict future glucose levels. These predictions enable diabetic patients and healthcare providers to take proactive measures to manage their blood sugar levels more effectively.

Objectives: The main objectives of this project are:

- Develop an accurate and effective model for predicting blood glucose levels using machine learning techniques.
- Improve the existing model by tuning parameters and improving the model architecture to increase accuracy.
- Expand data analysis by adding new analytical techniques that give users deeper insights into the causes of fluctuations in glucose levels.
- Improve the user interface by creating an interactive platform that displays glucose trends and forecasts more clearly and interactively.
- Provide treatment recommendations to diabetic patients based on the forecast results to help them better manage their condition.
- Evaluate the model's performance and compare it to previous work in this field.

To systematically explore the various aspects of this project, the report has been divided into several interconnected units that gradually explain the theoretical concepts, practical applications, and the steps that were followed.

Chapter 1: Introduction to Artificial Intelligence and Machine Learning in This chapter covers the basic principles of artificial intelligence and machine learning, including their definitions, types, applications, advantages and disadvantages. It also explains the relationship between machine learning and artificial intelligence, classifies learning types (supervised, unsupervised, reinforcement), and the importance of data quality in building effective models.

Chapter 2: Study of AI Applications in Glucose Level Predictions in This chapter focuses on type 1 diabetes and the role of artificial intelligence in predicting glucose levels. It also reviews the data sources used, database components, and key previous work in the field, highlighting the gaps the project seeks to fill.

Chapter 3: Study of Completed Work This chapter describes the practical work completed, starting with defining the problem, preparing data, testing several models (LSTM, FFNN, and SVM), adding and testing other models to analyze their performance using metrics such as RMSE and CEGA, and comparing all models. Furthermore, we developed a simple interface and provided treatment recommendations based on the predictions.

CHAPTER 1

Introduction to Artificial Intelligence and Machine Learning

1 Introduction

This chapter provides a comprehensive introduction to artificial intelligence and machine learning, two rapidly growing fields that significantly impact various aspects of life. This chapter explores what artificial intelligence is, its importance, definition, fields, languages and tools used in artificial intelligence, its advantages and disadvantages, and its expected future. This chapter aims to provide the reader with a basic understanding of these technologies and enable them to explore further knowledge about this exciting field.

In the modern era, artificial intelligence and machine learning are gaining increasing importance, experiencing rapid growth and significant impact on various industries.

For example, the artificial intelligence market size is expected to reach USD 243.72 billion in 2025, The market size is expected to exhibit a compound annual growth rate (CAGR) of 27.67% from 2025 to 2030, resulting in a market size of USD 826.73 billion by 2030 [1].

Artificial intelligence performs tasks that are usually performed by humans, such as patients and doctors, thus simplifying their lives at a lower cost.

AI is prominent in healthcare in many ways, such as tracking personal cases and predicting the development of chronic diseases, such as diabetes, as preventative measures, finding new links between genetic codes, operating robotic surgical assistants, automating administrative tasks, personalizing treatment options, and much more.

Although the two terms are sometimes confused, machine learning is one of the many branches of artificial intelligence. AI refers to any technique that gives a machine the ability to mimic human intelligence, while machine learning refers to the set of techniques and algorithms that allow machines to learn from data without being explicitly programmed.

2 artificial intelligence

2.1 What is artificial intelligence and its importance?

The concept of artificial intelligence (AI) is a new field that emerged around 1950 by Alan Turing (some call him the godfather of artificial intelligence). In short, it means making machines think and act like humans.

Initially, this field encountered many difficulties because scientists were trying to simulate human intelligence and thinking, trying to make machines think and act like humans. Because of this thinking, scientists failed to achieve real progress in this field. Despite all the scientific advances, scientists were unable to determine how humans think.

Scientists later came to the conclusion that we, as humans, don't care how a machine works (i.e., we won't restrict the machine to the way humans think). What matters is that we get the same results we get from humans. Humans see images as reflections of light, while machines see images as an array of pixels, zeros, and ones. At that time, artificial intelligence began to develop and entered all areas of our practical lives, whether military, medical, industrial, commercial, or even educational and entertainment. Performance accuracy was very high, and in some areas, the error rate was almost zero. In other words, machines began to mimic human thinking, even becoming better than humans in many areas. Today, science is the science of artificial intelligence. If the Internet was the revolution of the past years, then artificial intelligence is the revolution of the present and the future.[2]

It aims to make our lives easier and can be applied in almost all fields to achieve greater efficiency. This includes:

• Big Data Analysis: The amount of data currently available online far exceeds the ability of humans to comprehend, interpret, and make complex decisions based on it. Artificial intelligence algorithms can process, analyze, and understand this data, thus providing organizations with insights into their operations that they may not have previously been aware of.

• Decision Making: AI algorithms can sometimes make more accurate decisions than humans due to their ability to analyze complex and multiple relationships and leverage the vast amount of data available online.

2.2 Definition of Artificial Intelligence (AI)

According to Andreas Kaplan and Michael Heinlein, artificial intelligence can be defined as follows:

"The ability of a system to correctly interpret external data, learn from this data, and use that knowledge to achieve specific goals and tasks through flexible adaptation." [3].

The history of AI can be traced back to the 1950s, when this field saw significant developments thanks to the efforts of pioneers such as Alan Turing and John McCarthy. Turing introduced the famous Turing test to assess a machine's ability to exhibit intelligent behavior equivalent to that of humans. McCarthy coined the term "artificial intelligence" and founded the Artificial Intelligence Laboratory at MIT. [4]

AI technologies are categorized by their capacity to mimic human characteristics, the technology they use to do this, their real-world applications, and the theory of mind, which we'll discuss in more depth below.

Using these characteristics for reference, all artificial intelligence systems real and hypothetical fall into one of three types:

- **Artificial narrow intelligence (ANI):**
  which has a narrow range of abilities.
- **Artificial general intelligence (AGI):**
  which is on par with human capabilities.
- **Artificial superintelligence (ASI):**
  which is more capable than a human. [5]

2.3 Fields of Artificial Intelligence

Artificial intelligence is a broad field of study that encompasses many different theories, methods, and techniques, the most prominent of which are:

- **Machine learning:**
  Arthur Samuel defined machine learning as the field that gives computers the ability to learn from problems, they encounter without explicit instructions that is, the ability to tackle new problems.

- **Artificial neural networks:**

  A set of algorithms designed in a way inspired by neurons in the human brain, designed to recognize patterns.

- **Deep learning:**

  It uses large neural networks with multiple layers of processing units, leveraging significant computational advances (e.g., powerful processors) and improved training techniques to learn complex patterns from large amounts of data. [6]

2.4 The Four Types of Artificial Intelligence

In general, there are four types of artificial intelligence, which are:

- **Reactive AI**

This is the most basic type of AI. Reactive AI can operate based on an assessment of the current situation but is unable to build a store of memories for future use.

- **Limited Memory**

Limited memory AI can "remember" past experiences as pre-programmed representations of its environment. Limited memory AI then incorporates these memories into future decisions.

- **Theory of Mind**

This type of AI is more advanced than limited memory. Its name comes from the psychological term, ToM AI can attribute mental states such as beliefs, intentions, desires, emotions, and knowledge to others.

- **Self-Awareness**

Self-aware AI goes beyond ToM AI in that it has the ability to form representations of itself—and thus possesses consciousness.[7]

2.5 Artificial Intelligence Applications

Artificial intelligence is widespread in many fields and offers innovative solutions to various challenges. Its most important applications include:

- **Healthcare:**

o Medical Diagnosis: AI can analyze medical images such as X-rays and MRIs to accurately detect specific diseases

o Drug Discovery: AI is used to analyze genetic and biological data to develop new and effective drugs.

o Chronic Disease Tracking: such as predicting blood glucose levels.

o Providing Treatment Recommendations: Building an intelligent system that provides treatment advice based on the results of the predictions. These recommendations may include:

- ✓ Suggesting dietary modifications.
- ✓ Recommendations regarding insulin doses.
- ✓ Tips for improving physical activity.

- **E-Commerce:**

o Personalized Recommendations: AI is used to analyze customer behavior and provide personalized product recommendations.

o Automated Customer Service: AI-powered chatbots are used to provide instant answers to customer questions.

o Consumer Behavior Analysis: AI helps understand consumer needs and trends.

- **in Space:**

The role of artificial intelligence (AI) is to replace humans, as space exploration is a dangerous task for them. This is possible thanks to robots and expert systems that mimic human work. One example of this is the Mars rover, which can walk alone on the surface of Mars.

NASA is also working to use other AI technologies in space exploration, automating the analysis of images of galaxies and planets, developing autonomous smart spacecraft that can avoid space debris without human intervention, and creating more efficient and distortion-free communication networks using AI-powered devices. [8]

- **In digital marketing and advertising (Google's case):**

Google is increasingly relying on tools capable of automatically suggesting, creating, and delivering ads, increasing efficiency and reducing costs, such as Performance Max.

- o Impact:
- ✓ This has led to a restructuring of its 30,000-employee sales department.
- ✓ Google believes many sales roles are becoming redundant.

- ✓ Artificial intelligence is not just used as an assistant, but has become an actual replacement for some human roles. [9]

2.6 Advantages of Artificial Intelligence

Artificial intelligence offers countless advantages, a few of which are:

- Makes machines more powerful and useful.
- Offers new approaches to problem solving.
- Is better at processing information than humans.
- Improves work efficiency, reducing the time required to complete a task compared to humans.
- Often more accurate than humans.

2.7 Disadvantages of Artificial Intelligence

On the other hand, there are many disadvantages, including:

- The inability to generalize from one task to another. That is, a machine can only perform a specific task (or several tasks) for which it has been previously trained, and cannot perform a task for which it has not been previously trained.
- Cost (the cost of implementing AI applications is extremely high).
- Lack of skills (there are few skilled programmers capable of developing AI programs).
- Requires deep technical expertise.
- Robotics is one of the applications of AI that is replacing jobs held by humans, and thus may lead to increased unemployment.

3 Machine Learning

3.1 Definition of Machine Learning

Professor Tom Mitchell, an American computer scientist and professor at Carnegie Mellon University in the United States, defined machine learning as follows:

"When a computer program is instructed to learn from experience E to perform a task T, the accuracy of which is measured by the accuracy metric P. If the program can perform the task T with the accuracy specified by P by learning from experience E, then that is machine learning."

Machine learning models are nothing more than algorithms built on the principles of linear algebra, calculus, and statistics. Therefore, these branches of mathematics are important for those studying machine learning to better understand what is happening behind events and what the algorithms do. Of course, trying to understand why these algorithms work can be very difficult, and that is the specialty of the artificial intelligence researchers who invent and develop these algorithms. As for us programmers, all we want is to understand how these algorithms work. Machine learning has a significant impact on other scientific fields. It is used to analyze DNA sequences, understand stars, find distant planets, diagnose diseases, develop medicines, and much more. Machine learning is a subspecialty of artificial intelligence, but even machine learning is divided into other subspecialties, such as deep learning, which specializes in solving complex problems and applications by simulating a neural network from the human brain. This branch works to solve problems such as facial recognition applications in images and others. [10]

The following image illustrates the relationship between artificial intelligence, machine learning, and deep learning.

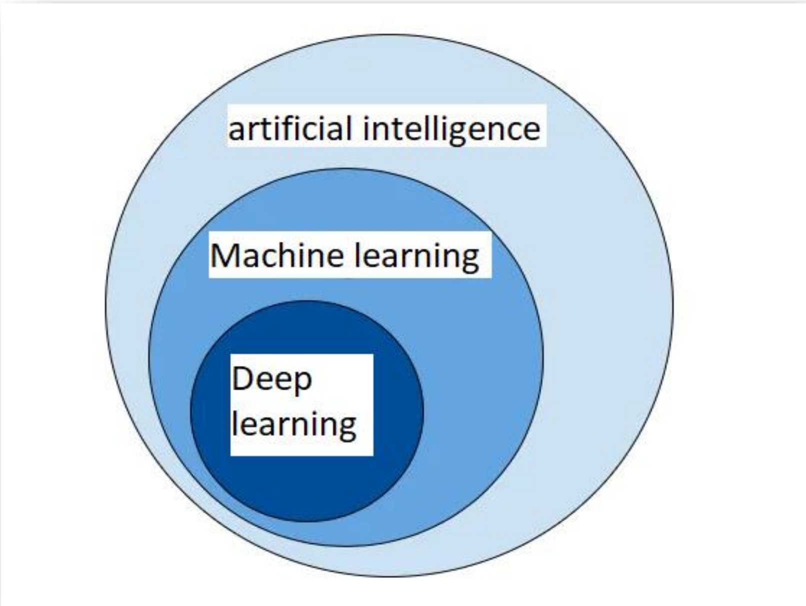

Figure 1: the relationship between Artificial Intelligence, Machine Learning, Deep Learning

Machine learning relies heavily on big data and the quality of that data. Therefore, machine learning overlaps with another discipline called data science. The boundaries between the two are not clear, but we can say that machine learning relies on data science, not the other way around. A data scientist can work with big data and improve its quality using pre-studied methods until the data is truly ready to train the machine learning model. This is usually done by a machine learning engineer. In many cases, especially in small companies, there is no data science specialist to improve the data before sending the data to a machine learning engineer to train the model. It's okay if things seem unclear to you so far.

In the following image we see the relationship between artificial intelligence, machine learning, deep learning, and data science.

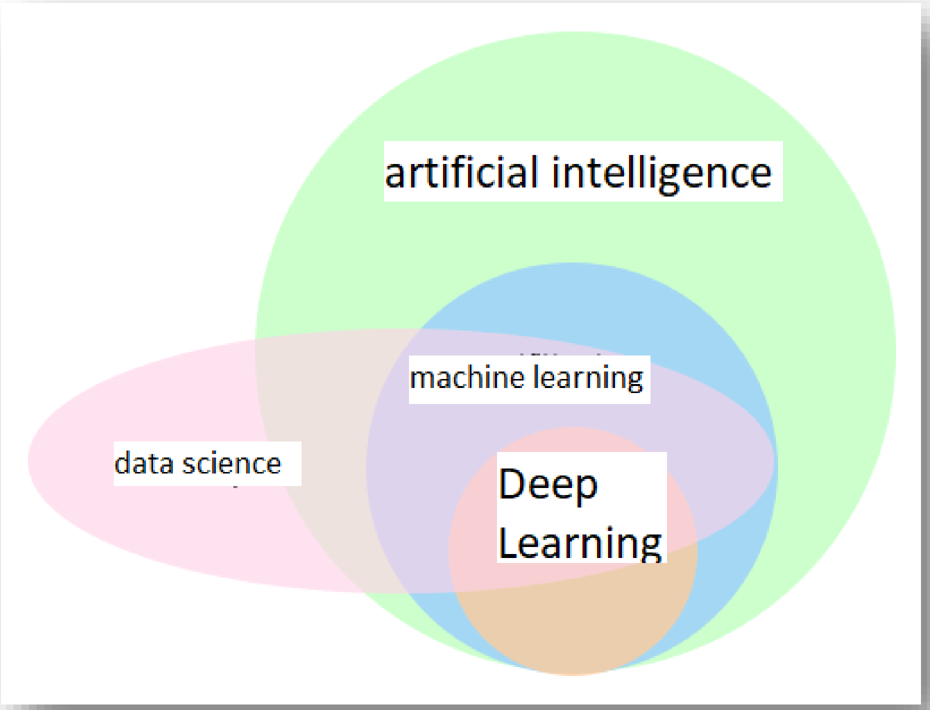

Figure 2: the relationship between Artificial Intelligence, Machine Learning, and Deep Learning, and data science.

Basic machine learning algorithms include linear regression, which is used to model the relationship between two or more variables, and decision trees, which are used to make decisions based on a series of rules.

3.2 Types of Machine Learning

Machine learning algorithms can be classified into three main types:

- **Supervised Learning:**

  In this type, the model is provided with labeled data that includes the desired inputs and outputs.

Examples of common supervised learning algorithms include neural networks and support vector machines

- **Classification**

Classification is the process of assigning inputs to predefined classes. This method is used in applications such as classifying emails as spam or otherwise.

The most common classification algorithms are:

- Support Vector Machines (SVM)
- Random Forests (RF)
- Decision Trees (DT)
- Logistic Regression

- **Regression**

Regression is used when the outputs are continuous numbers. An example is predicting house prices or blood sugar levels.

The most common regression algorithms:

- Linear Regression
- Polynomial Regression
- Support Vector Regression (SVR)

- **Unsupervised Learning:**

  In unsupervised learning, the model is trained on unlabeled data, meaning it has no known outputs. This type of learning is used to discover patterns in data or reduce dimensionality.

- **Clustering**

Clustering is used to divide data into groups with similar characteristics. This method is applied in applications such as segmenting customers in a market.

The most common clustering algorithms:

- K-Means Clustering
- Hierarchical Clustering
- DBSCAN

- **Dimensionality Reduction**

Dimensionality reduction is used to simplify data while preserving important information. This technique helps improve model performance and reduce processing time. The most popular dimensionality reduction techniques:

- Principal Component Analysis (PCA)
- t-SNE

- **Reinforcement Learning**

Reinforcement learning is an approach based on a system's interaction with a specific environment, where the system receives rewards for making correct decisions and penalties for making incorrect ones. Reinforcement learning is used in applications such as games, robotics, and self-driving cars.

The most popular reinforcement learning algorithms:

- Q-Learning
- Deep Q-Network (DQN)
- Proximal Policy Optimization (PPO)

The following image shows types of machine learning

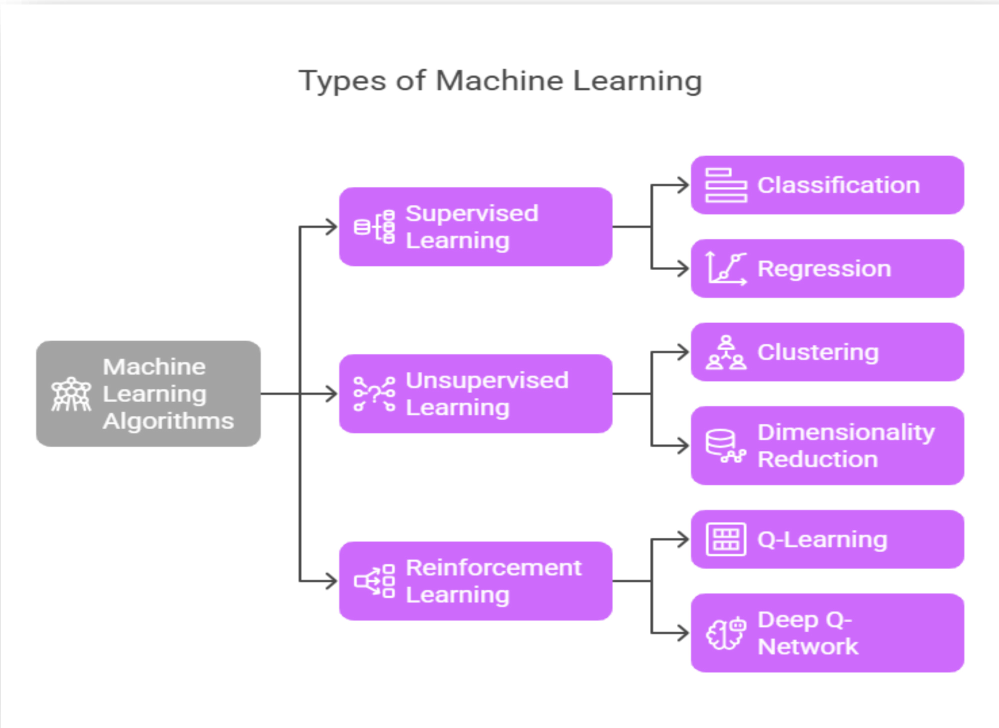

Figure 3: types of machine learning.

3.3 Comparison of learning types

To understand the fundamental differences, we can compare the types of learning in artificial intelligence in terms of several basic aspects, as shown in the following table:

Learning Type	**Inputs**	**Outputs**	**Examples of Use**
Supervised Learning	Labeled Data	Known	Image classification, stock price prediction
Unsupervised Learning	Unlabeled Data	Unknown	Pattern discovery, customer behavior analysis
Reinforcement Learning	Interactive Environment	Decision that improve performance	robots, video games

Table 1: Comparison of learning types

4 The Importance of Understanding and Improving Data

Machine learning, with all its models and algorithms, relies primarily on data, regardless of whether it is both input and output, or even just input. Therefore, data quality and understanding are of paramount importance. Sometimes you will need to add or remove a feature from the data, which is called feature engineering or feature extraction. Other times, you will need to combine features in some way. All of these things are found in the discipline of statistical learning, which was previously referred to. Understanding, purifying, and improving the data is one of the most important stages. This allows you, as a machine learning engineer, to choose the appropriate algorithm for the data from the available machine learning algorithms that will provide the best possible accuracy based on the task you want to perform with the model you have programmed. Data quality also plays a significant role in the accuracy rate of the model. By improving data quality, we can achieve 99% accuracy instead of 97%, and we can achieve 97% accuracy instead of 80% by modifying the values of the algorithm parameters. Therefore, understanding the data is the first step towards selecting the algorithm and parameter values. The programming part of machine learning models is just one point in a bigger picture, which is the problem to be solved, so it is very important to always keep that big picture in mind when programming a model. Based on that big picture, you will be able to determine whether you have the right data or not, and that is the first step. Then, after that, you will be able to determine what to do during the feature engineering phase by purifying and improving the quality of the data. Then, based on that picture, you will be able to determine the correct algorithm for the application you are working on.

5 Conclusion

In this chapter, we have provided a comprehensive introduction to artificial intelligence and machine learning, from their basic definitions to their broad applications. We have seen how AI is experiencing rapid growth and profound impact on various fields, especially medicine, which we will discuss in detail later, and how it is expected to play an increasing role in solving global problems and improving our lives.

It is essential to understand artificial intelligence and machine learning and their impact on the future. We must invest in the development of artificial intelligence responsibly and ethically, ensuring that these technologies are used to improve the lives of all humanity.

Through continuous learning, we can all contribute to shaping the future of artificial intelligence.

I invite you to explore further knowledge about this exciting and ever-evolving field, especially in the health aspect, which I will address in this project, striving to develop meaningful and engaging work.

CHAPTER 2

Study of AI Applications in Glucose Level Prediction

1 Introduction:

In diabetes, blood glucose cannot enter cells and accumulates in the bloodstream, leading to high blood sugar levels over time. High blood sugar damages the body and can lead to diabetes-related complications if left untreated. In type 1 diabetes, the immune system mistakenly treats the beta cells in the pancreas, which are responsible for producing insulin, as foreign invaders and destroys them. When enough beta cells are destroyed, the pancreas cannot produce insulin or produces too little, requiring insulin to survive. Insulin is a hormone that helps blood glucose (blood sugar) enter the body's cells for use as an energy source. [11]

Machine learning can be used to predict blood sugar levels. Predicting blood glucose levels is a powerful tool in developing personalized decision-making systems and low- and high-blood sugar alarms.

It is best for patients to monitor themselves regularly and systematically. This is what current treatment approaches rely on to determine how to manage their disease, keep blood sugar levels within the recommended range, and prevent serious complications that may arise.

This means effectively monitoring glucose levels, injecting and dosing insulin, and managing diet and physical activity. Recent advances in diabetes management technologies and wearable devices have made collecting relevant data easier. We believe that machine learning (ML) techniques can leverage the available data and help patients manage their blood sugar levels.

After extensive research on this topic, I came across an excellent piece by Mr. Ladislav on GitHub. The main motivation for choosing this topic is that the author of this thesis has type 1 diabetes and understands the complexity of managing this disease.

A blood glucose alarm, which alerts a patient to high or low glucose levels, could have a significant impact on type 1 diabetes management. If the model inputs include information about insulin and diet, it could help the patient make treatment decisions, suggesting the appropriate insulin dose or carbohydrate intake to maintain blood glucose levels within a safe range.

2 Diabetes:

2.1 Definition of diabetes

Diabetes is a chronic, metabolic disease characterized by elevated levels of blood glucose (or blood sugar), which leads over time to serious damage to the heart, blood vessels, eyes, kidneys and nerves. The most common is type 2 diabetes, usually in adults, which occurs when the body becomes resistant to insulin or doesn't make enough insulin. In the past 3 decades the prevalence of type 2 diabetes has risen dramatically in countries of all income levels. Type 1 diabetes, once known as juvenile diabetes or insulin-dependent diabetes, is a chronic condition in which the pancreas produces little or no insulin by itself. For people living with diabetes, access to affordable treatment, including insulin, is critical to their survival. There is a globally agreed target to halt the rise in diabetes and obesity by 2025.

About 830 million people worldwide have diabetes, the majority living in low-and middle-income countries. More than half of people living with diabetes are not receiving treatment. Both the number of people with diabetes and the number of people with untreated diabetes have been steadily increasing over the past decades. [12]

2.2 Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes 2021:

Diabetes can be classified into the following general categories:

1. Type 1 diabetes (due to autoimmune β-cell destruction, usually leading to absolute insulin deficiency, including latent autoimmune diabetes of adulthood)
2. Type 2 diabetes (due to a progressive loss of adequate β-cell insulin secretion frequently on the background of insulin resistance)
3. Specific types of diabetes due to other causes, e.g., monogenic diabetes syndromes (such as neonatal diabetes and maturity-onset diabetes of the young), diseases of the exocrine pancreas (such as cystic fibrosis and pancreatitis), and drug- or chemical-induced diabetes (such as with glucocorticoid use, in the treatment of HIV/AIDS, or after organ transplantation)
4. Gestational diabetes mellitus (diabetes diagnosed in the second or third trimester of pregnancy that was not clearly overt diabetes prior to gestation)

This section reviews most common forms of diabetes but is not comprehensive. [13]

2.3 Symptoms:

Symptoms include:

- ✓ Urinating often
- ✓ Feeling very thirsty
- ✓ Feeling very hungry—even though you are eating
- ✓ Extreme fatigue
- ✓ Blurry vision
- ✓ Cuts/bruises that are slow to heal
- ✓ Weight loss—even though you are eating more

It's important to know when you first develop type 1 diabetes, you may not have any symptoms at all. [14]

2.4 Insulin:

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

Insulin pen is a reusable syringe with smart controls. Some pens have features like a memory mechanism where the pen can display the time and dosage of the most recent insulin injection. Patients usually have a set of 2 pens, one containing rapid-acting insulin and the other a long-acting insulin.

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

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

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

The patient providing data for this thesis uses insulin pens, therefore our data contains records of rapid and long insulin. [15]

2.5 BG Dynamics:

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

The main factor that makes BG rise is carbohydrate consumption. After consumption, car bohydrates are broken down in the body into glucose, a very simple carbohydrate that the body uses as fuel for various physiological processes. The time between meal consumption and glucose availability in the blood varies between different foods. The length and rate of glucose release after food consumption for various foods is classified using Glycemic Index (GI) tables [16]. Some other factors that make BG rise are stress, anaerobic exercise, effects of glucagon– a hormone that stimulates the release of glucose stored in the liver, and others.

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

It should also be noted that BG response can change depending on the daily insulin dose, as Davidson et al. demonstrated with adult patients using rapid insulin [17]. This effect can be summed up as follows: the higher the overall insulin dose taken daily by the patient, the lower the BG sensitivity to insulin. Furthermore, BG dynamics will vary between patients, for example, based on the body weight, as was again demonstrated by Davidson et al. [17]

3 Dataset:

When conducting the study and examining previous work on predicting blood glucose levels, we noticed that they relied on medical databases collected from reliable official sources, such as OhioT1DM [18], as well as databases published on research platforms such as IEEE [19] and Kaggle. [20]

3.1 common points:

These datasets have which make them suitable for training artificial intelligence models. These are as follows:

1. Type: Time series data recorded periodically at regular intervals (such as every 5 or 15 minutes).
2. Source: Data collected using continuous glucose monitoring (CGM) devices such as the Dexcom G5 or FreeStyle Libre.
3. Reliance on real data collected from patients with type 1 and type 2 diabetes using CGM devices.
4. Multidimensional data: In addition to glucose measurements, it also includes insulin doses, meal times, physical activity, and even physiological indicators such as heart rate and temperature.
5. Standard data formats such as CSV to facilitate their use in deep learning and time series analysis environments.

The data are often They are personalized or individual (from a single patient or a limited number of patients) and require preprocessing such as cleaning, time standardization, and handling missing values. In some cases, the data is partial or lacks certain information, such as physical activity or sleep.

3.2 Their primary purpose:

is to train the results of artificial intelligence models such as LSTM, GRU, CNN, or LMU-RNN to:

• Predict future glucose levels
• Assess the risk of hypo- or hyperglycemia
• Provide treatment recommendations or proactive alerts

These databases have a common research goal: to support the development of accurate and effective predictive models for diabetes management and improving patients' quality of life.

3.3 Common components of the dataset:

The common components of the dataset used in previous blood glucose prediction studies:

Data type	Description
Glucose levels	Blood glucose values record throughout the day using CGM devices
Insulin intake	Dosage of insulin administrated (both rapid-acting and long-acting)
Carbohydrate intake	Amount of carbohydrates consumed during meals
Physical activity	Information about exercise or physical activity (included in some datasets)
Timestamps	The exact time for each reading or event
Manual annotations	Optional notes such as symptoms, sleep, or unusual events (in some datasets)

Table 2: Common components of the dataset.

4 Previous work:

Recent years have witnessed increased interest in models for predicting diabetes levels worldwide, in an effort to improve the lives of diabetes patients and help them develop more effective treatment strategies and approaches. Research in this field is abundant, but effective applications are unlimited and face technical and practical challenges that hinder their widespread adoption. Among the most prominent research tasks that can be highlighted is DeepCGM [21], which expands on Convolutional Networks (CNNs). This model is characterized by its speed and ability to train for long-term planning, but it suffers from performance fluctuations when attempting longer-term predictions, in addition to the need for precise parameter tuning to obtain effective results.

However, the Diaredict [22] project has turned to using advanced algorithms such as GRU and Transformer to analyze different data and provide analytical experts based on the analyses. Despite the outstanding results, the risk of the adopted model being slow and inappropriate, and the training remains too large to reach the excellent level of data accuracy.

Another noteworthy project is Glucose [23], which combines the capabilities of LSTM and GRU to improve long-term performance while providing an interactive graphical interface that makes it easier for users to understand. However, the absence of symptoms is notable. The service is limited, but the network may be a barrier for some unregistered users.

Another recent project is predicting blood glucose levels using LMU-RNNs [24], which are popular on Legendre memory units. This allows them to handle long-term data, but they require computational resources and can be implemented on devices with limited resources. Finally, there is a blood glucose prediction project [25] based on LSTM and model data from nearby sensors (CGM). This model is excellent for its short-term accuracy and is one of the small-scale applications that can be easily built and developed.

5 Conclusion:

After learning about the disease, its severity, and the challenges associated with its treatment, we find that type 1 diabetes patients face daily challenges. Therefore, modern technologies such as machine learning must be relied upon to control the disease. Predicting blood glucose levels not only aims to prevent episodes of high or low blood sugar, but also serves as a personal assistance tool that enhances the patient's ability to make accurate and rapid treatment decisions. Based on real data collected from continuous monitoring devices, deep learning models can be built that provide immediate and effective recommendations and contribute to reducing the daily mental burden borne by the patient. Therefore, integrating these models into smart systems or health applications could transform diabetes management. This project, based on personal experience and a deep understanding of the disease's details, aims to develop an effective predictive model that helps patients improve their quality of daily life and represents a step toward smarter and more humane treatment decision support systems.

CHAPTER 3

Practical Implementation, Optimization, and User-Centric Integration

1 Introduction:

We previously discussed some basic facts about type 1 diabetes. In this module, we delve deeper into the topic, explaining the nature of the condition in greater detail, discussing the challenges associated with regulating blood sugar and its daily dynamics, and finally presenting and analyzing the work accomplished within this project.

This module focuses primarily on the practical and technical aspects of the blood glucose prediction project. The project relied on integrating multiple data sources, such as continuous glucose (CGM) measurements, insulin doses, carbohydrate intake, and physical activity indicators. The primary goal of type 1 diabetes treatment is to maintain glucose levels within a safe range, which requires the patient to closely monitor their condition through daily self-management, including tracking food, physical activity, and insulin doses.

With the increasing use of continuous glucose monitors (CGMs) and biometric tracking devices such as smartwatches and fitness devices, it has become possible to collect accurate and comprehensive data about a patient's daily life. This data represents a real opportunity to build predictive models that help patients make better, more informed decisions. When I first started working on the project, I was looking for a practical way to get started without duplicating what had already been done. I found that the best starting point was to leverage one of the open-source projects that addressed the same problem. Among the projects I reviewed, one titled "Blood-Glucose-Prediction", posted on GitHub by a user known as MrBlueHere, caught my attention. I liked its organization and its reliance on an LSTM model to predict blood glucose levels, which provided me with a ready-made foundation from which to build and develop.

Although the original project provided a reasonably efficient prototype, I found that there were many areas that could be improved. Therefore, I set myself three main goals during this work:

- First, to improve the accuracy of the original model by optimizing data preprocessing.

- Second, Experiment with several other classical and modern models that are more accurate in processing time series data in order to compare and discover the appropriate model for this work.

- Third, provide recommendations based on the results of the forecasts.

- Finally, I have added an easy to use interactive interface for any simple user to use, in which data is uploaded, results are displayed, and recommendations are made.

In academic projects such as the capstone project, this type of work involves adding improvements to the code, developing it, and achieving scientific reproducibility, which saves time and effort during experiments. It also reflects a programming awareness of fundamental software engineering principles, most notably the DRY (Don't Repeat Yourself) principle, and an emphasis on building an organized software structure that facilitates subsequent maintenance and development, based on the belief that good code must not only be efficient but must also be reusable and extensible.

2 Problem description

Despite significant advances in diabetes monitoring technologies, type 1 diabetes patients still face a persistent challenge in accurately predicting glucose level fluctuations. Rapid and unpredictable changes in blood sugar can be caused by several interrelated factors, such as meal timing, carbohydrate intake, physical activity, or even stress and illness. These fluctuations are often difficult for patients to manually predict or proactively manage. This difficulty represents a clear weakness in traditional self-management systems and opens the door to leveraging AI technologies to build data-driven tools that provide accurate and reliable predictions.

Accordingly, the problem of this project was posed as follows:

Is it possible to build an intelligent model capable of accurately predicting blood glucose levels based on a patient's daily data, thus contributing to reducing episodes of sudden highs or lows and supporting patients in making proactive decisions to better manage their condition?

3 Understanding some basics

3.1 Type 1 Diabetes and BG Dynamics [15]

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

The main purpose of T1D treatment is to keep the BG in a safe range, by that we understand the observed range of a healthy person. The term for this safe range is euglycemia. The exact numbers used to classify euglycemia vary across the academic literature, but generally are around 60 mg/dl – 140 mg/dl or 3,3 – 7,8 mmol/l. This range is commonly used for diabetes tests [26].

Hypoglycemia is a condition when BG levels drop below 3,3 mmol/l or 60 mg/dl. Symptoms of hypoglycemia range from anxiety, sweating, and hunger, to neurological impairments, including behavioral changes, cognitive dysfunction, and in extreme cases seizures, and coma [27]. Generally, the lower the glucose, the higher the severity of symptoms. Patients' objective is to avoid hypoglycemia, as it impairs their function at best, and threatens their life at worst.

Hyperglycemia is a condition when BG is above 180 mg/dl or 10 mmol/l. Chronic levels of BG above this threshold can produce noticeable organ damage over time. Symptoms of hyperglycemia are usually benign like dry mouth and polyuria but can be more severe if hyperglycemia develops into ketoacidosis. The greatest danger of chronic hyperglycemia, which is frequent and prolonged incidents of hyperglycemia are the long-term effects, especially on the microvascular system. These effects may be life-threatening and include damage to the eye, kidneys, nerves, heart, and the peripheral vascular system [28].

3.2 BG Prediction [15]

It was shown by Martin H Kroll [29] that BG variation includes a deterministic component and that one can describe it as a nonlinear oscillatory or chaotic system, that can be modeled.

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

Since BG dynamics are influenced by factors like insulin, physical activity, and others, it makes sense to add these features to give the model more context. It is true that to a certain degree this information is already captured in the BG levels, but it may not provide the full picture.

Therefore, the prediction task is extended. The goal is to find a function f (x1...,xn) which takes an ordered set of vectors as an input and its output is the BG at time t+PH. Each input vector xi is associated with one of the features (BG, insulin, carbohydrates, ...). Each of the input vectors contains a finite number of elements and in the simple form, it will be the s successive previous measurements. However, greater input customizability is supported, by allowing the window of historical measurement to be variable in size between individual feature vectors. More than that, feature vectors that are transformations of the previous measurements are also permitted. The transformation may be the sum, mean, or other function applied to previous values.

It is expected that with a longer PH, the prediction performance will decrease. The uncertainty increases with the amount of time passed and so does the domain of possible BG values that come with a larger PH. It takes time for BG to change, and as PH increases, the BG has further opportunity to develop, thus the domain of possible BG values increases too.

3.3 A brief and accurate description of the work in the original project

In this project, the author focused on developing models capable of predicting blood glucose levels in a patient with type 1 diabetes, using artificial intelligence and machine learning techniques. Data was collected from a single patient over 128 days of normal daily life, including continuous glucose (CGM) measurements, insulin doses, carbohydrate intake, and physical activity recorded by a fitness tracker (Fitbit).

- **Data Processing and Analysis**

Different data sources were combined, and multivariate time series with a 15-minute interval were generated.

The author used simplified models to estimate insulin activity in the body, including the "insulin-on-board" model inspired by the LoopKit system, to calculate the amount of active insulin during the prediction intervals.

Additional features such as "RapidInsulin6d" and "RIOB" were created to improve prediction accuracy.

- **Model Building and Evaluation**

Three types of models were used: recurrent neural networks (LSTM), forward-facing networks (FFNN), and support machine (SVM).

The models underwent parameter tuning using techniques such as HyperBand with early stopping to avoid overlearning.

Performance was evaluated using two key metrics:

RMSE to measure numerical prediction accuracy

CEGA to assess clinical accuracy.

- **Results**

The LSTM model achieved the best results at prediction horizons of 30 minutes and 60 minutes, while the FFNN model outperformed at prediction horizons of 120 minutes.

SVM models demonstrated the best clinical accuracy across different prediction horizons, with the largest proportion of results classified within Region A in the CEGA analysis.

Using the RIOB feature instead of raw insulin data was shown to improve performance by approximately 3%.

4 Implementation

4.1 Model Implementation Steps

Specifying the number of time steps (Window Size).

Adjusting the number of neural units within the layers.

Choosing the appropriate activation function (ReLU or Tanh).

Using the Adam learning algorithm.

Setting the number of iterations (Epochs) and batch size.

4.2 Environment and tools used

The project was implemented using Python, because it is one of the most widely used languages in the field of artificial intelligence and is easy to learn and use. Python provides a large number of useful libraries that helped me greatly in handling data, building models, and analyzing results. Among the most important libraries I used are:

TensorFlow: I used it to build an LSTM model because it is very suitable for temporal data.

Keras: A simple interface within TensorFlow that made it easy to design and test the model.

SciKit-learn: I used it to build an SVM model, as well as to partition data and evaluate performance.

Pandas and NumPy: These were essential tools for reading, organizing, and analyzing data.

Matplotlib and Seaborn: I used them to plot graphs that illustrate the results and data behavior.

As for the work environment, I used Jupyter Notebook because it allows me to write and run code step by step, with results displayed directly. This made it much easier for me to monitor and modify the project while implementing it.

5 Dataset

This project relied on three main datasets, collected from a single patient with type 1 diabetes over a relatively long period of 128 days. This data aimed to represent the patient's daily life under realistic conditions, allowing the development of a model that addresses multiple scenarios that represent actual challenges in disease management.

The first data source was a FreeStyle Libre glucose meter, which uses non-continuous real-time monitoring technology. The device measures the concentration of glucose in the interstitial fluid every 15 minutes, and these readings are stored internally until they are transmitted via NFC using a dedicated reader or smartphone. The data extracted from this device formed the backbone of the model, with values converted from mmol/L to mg/dL to standardize the format used for prediction.

The second data source was a personal electronic record documented by the mySugr mobile application. Through this application, the patient entered daily details, including insulin doses (both rapid and long-acting), carbohydrate intake, and the glycemic index of each meal. Although the recording process was manual, it provided important information used to estimate the expected impact of each meal or dose on blood sugar later.

The third source was a Fitbit Charge 4 activity tracker, which provided real-time data on heart rate, steps taken, distance traveled, and calories consumed. This data allowed the prediction model to add a dynamic and physiological dimension, enhancing its ability to understand the impact of physical activity on blood glucose levels.

By combining these three sources, a rich, multidimensional database was built, which was a key factor in effectively training the predictive models.

- The meter data is taken from text (.txt) files.
- The mySugr app data is taken from a CSV export.
- The Fitbit data is taken from JSON files, which are often exported via the user's Fitbit account.

- ### Integrating Data Sources into a Unified Time Series

After collecting data from the three different devices, the critical step was to unify them into a single, consistent time series that the learning model could handle effectively. Each of these sources relied on a different measurement frequency, which required careful reorganization and formatting of the data. A fixed time period of 15 minutes was chosen as the reference unit for collecting all data, as this represents the interval between CGM readings, the primary source of information we were trying to predict. The insulin and carbohydrate data were then matched to the closest time interval within this time frame, so that each dose or meal was recorded within its closest actual time window.

The data from the Fitbit, which is often recorded every minute or even every 5 seconds, was aggregated within each 15-minute window by calculating the average or sum of values based on the variable type (such as average heart rate or total calories burned).

This way, all daily events were transformed into a regular, structured timeline, with each time window representing a single record containing the following values:

Current glucose value

Carbohydrate intake

Insulin dose (rapid and long-acting)

Physical activity indicators

This integration enabled the model to see the full picture of the patient's condition at each moment, allowing it to learn from the entire temporal context rather than relying on a single, isolated value.

```
5]: glucose_df_resampled
```

5]:

	Time	Glucose	Rapid Insulin	Long Insulin	Carbohydrates	Glycemic Load	bpm	distance	calories	Hour	Rapid Insulin 6d
0	2021-12-01 00:00:00	7.3	0.0	15.0	0.0	0.00	60.312381	22.9	19.59	0.0	23.166667
1	2021-12-01 00:15:00	8.0	0.0	15.0	0.0	0.00	63.833466	73.5	17.96	0.0	23.166667
2	2021-12-01 00:30:00	8.6	0.0	15.0	0.0	0.00	51.559524	5.5	17.86	0.0	23.166667
3	2021-12-01 00:45:00	8.8	0.0	15.0	0.0	0.00	54.447937	0.0	17.62	0.0	23.166667
4	2021-12-01 01:00:00	9.1	0.0	15.0	0.0	0.00	51.072959	0.0	21.17	1.0	23.166667
...	...	...	...	...	...	...	...	...	...	...	...
12284	2022-04-07 23:00:00	8.6	0.0	9.0	5.0	2.25	72.603725	40.1	56.75	23.0	24.500000
12285	2022-04-07 23:15:00	8.6	0.0	9.0	0.0	0.00	70.983362	17.2	52.45	23.0	24.500000
12286	2022-04-07 23:30:00	8.4	0.0	9.0	0.0	0.00	78.193862	155.5	60.42	23.0	24.500000
12287	2022-04-07 23:45:00	8.9	0.0	9.0	0.0	0.00	84.487015	255.4	60.95	23.0	24.500000
12288	2022-04-08 00:00:00	NaN	0.0	9.0	0.0	0.00	97.285714	11.8	6.46	0.0	24.500000

12289 rows × 11 columns

Figure 4: Columns and rows used in data.

- Time: The timestamp for each recording.
- Rapid Insulin: Rapid-acting insulin doses.
- Long Insulin: Long-acting insulin doses.
- Carbohydrates: The amount of carbohydrates consumed.
- GI: Food description (often used to determine the glycemic index of foods).
- Body weight: Body weight (may help analyze body changes with diet).
- Fitbit dataset: Provides information about physical activity and heart rate.

6 Optimization:

6.1 Interpolation:

Handling missing values (Glucose, Insulin, Carbohydrates) using interpolation or smart deletion.

Improving model performance: Interpolation can improve model performance, especially if there are missing data that affect the model's accuracy. Filling in missing values can help achieve a more stable model.

Mitigating missingness: When there are missing values, especially in key columns like "Glucose," deleting these rows can result in significant data loss. Therefore, it may be better to fill in the missing values rather than deleting them.

Keeping up with modern developments: Using modern interpolation methods, such as linear or polynomial, can be a good option for improving the accuracy of predictions and making the model better at handling incomplete data.

The data is divided into three consecutive time series without overlap or gaps:

Training set: from 12/01/2021 to 28/02/2022

Validation set: from 28/02/2022 to 26/03/2022

Test set: from 26/03/2022 to 7/04/2022

This improvement in data quality represents an essential step towards ensuring a more accurate and reliable model in the next phases.

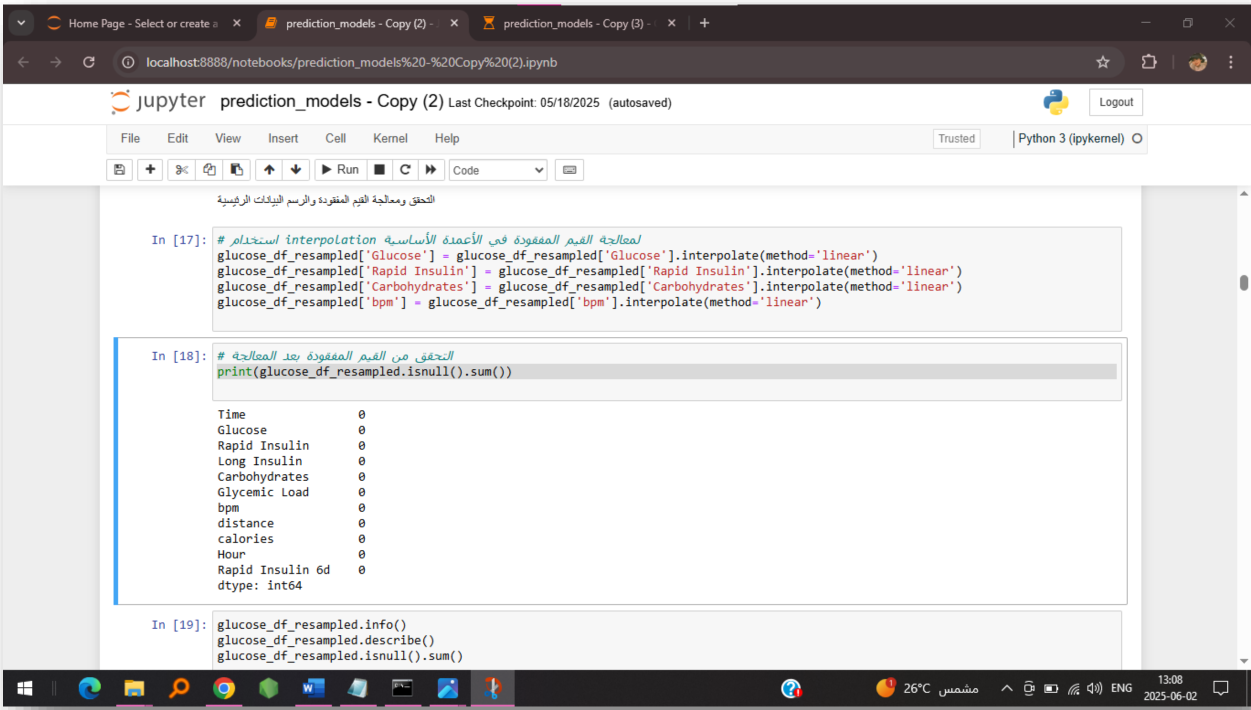

Figure 5: Data after processing and displaying missing values.

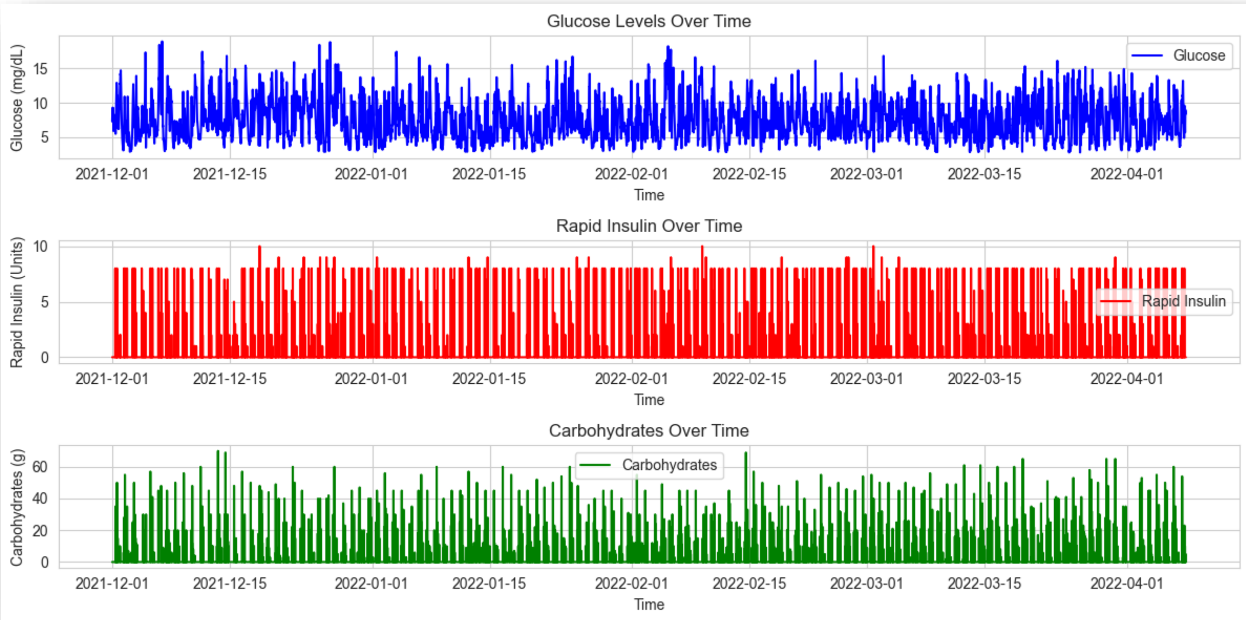

Figure 6: Data plot after processing.

7 Models implementation

7.1 Traditional regression models for predicting glucose levels:

Three traditional models were used to predict blood glucose levels: linear regression, second-order polynomial regression, and AdaBoost model boosted with decision trees. These models were used to evaluate their ability to predict glucose levels at multiple time points: 30 minutes, 60 minutes, 120 minutes, and more than two hours after the last available reading.

- **Data preparation:**

Time lags were generated based on previous glucose levels for up to three time points. The data was then split 70% for training and 30% for testing.

For each time point, the model was trained on its own data to ensure prediction accuracy.

- **Linear regression:**

Performance is good over short periods, but gradually deteriorates as the time period increases.

Over a short period (30 minutes):

RMSE = 0.5960

Train R^2 = 0.9688, RMSE = 0.5077

Over longer periods (60-120 minutes): Signs of deteriorating accuracy are evident.

- **Polynomial regression (Grade 2):**

It performed similarly to linear regression, with a slight improvement due to its ability to capture some nonlinear patterns.

For the 30-minute period:

RMSE = 0.5939

Train R^2 = 0.9692, RMSE = 0.5043

As the time period increased, it maintained relatively stable performance compared to the linear model, but accuracy declined after 120 minutes.

- **Prediction of Linear& Polynomial**

Below is the forecast for three consecutive periods 30 minutes, 60 minutes, 120 minutes:

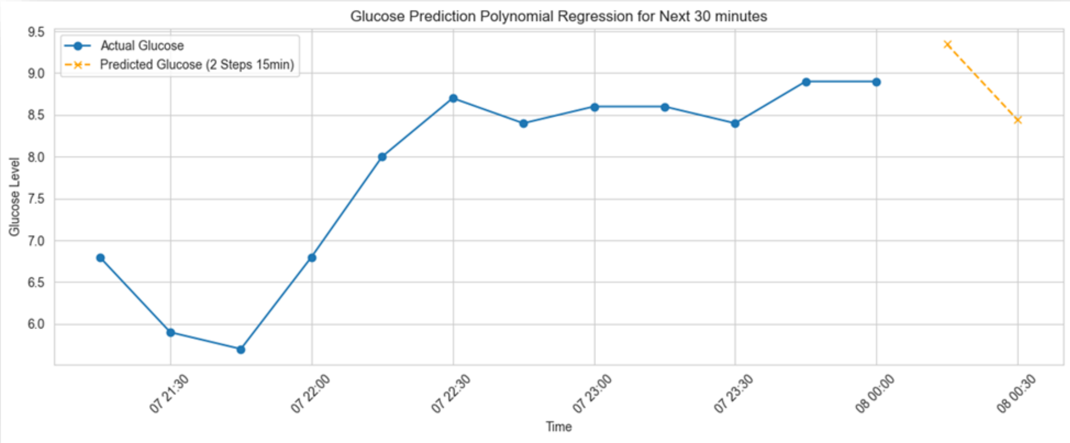

Figure 7: Prediction linear and polynomial regression for next 30 minutes.

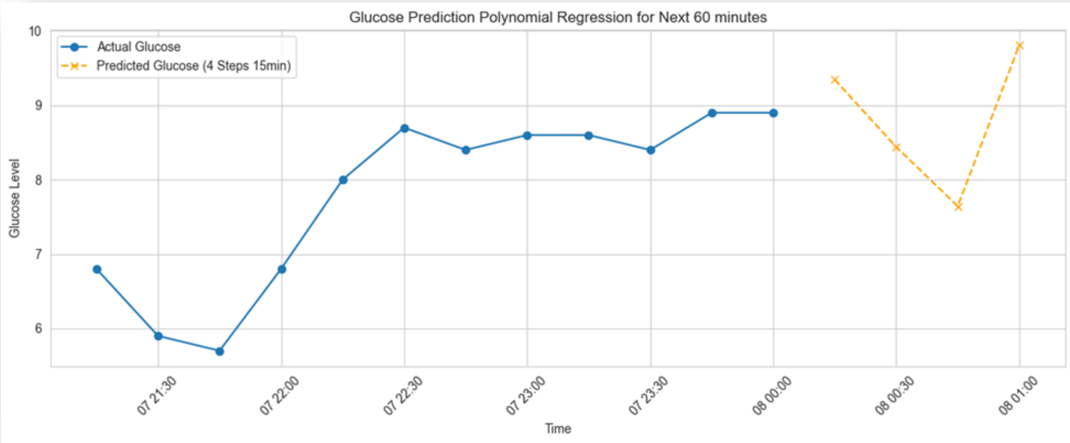

Figure 8: Prediction linear and polynomial regression for next 60 minutes.

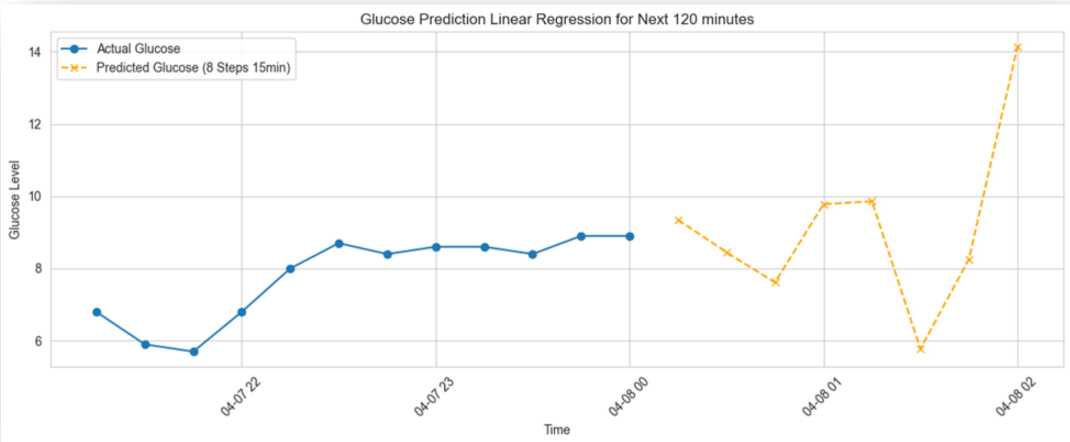

Figure 9: Prediction linear and polynomial regression for next 120 minutes.

- **AdaBoost:**

A more complex model, based on boosting the performance of decision trees, was tuned using 100 trees with a maximum depth of 4 and a learning level of 0.5.

For a 30-minute period:

RMSE = 0.7750

In short forecasts, it showed more volatility than other models, possibly due to its sensitivity to the data distribution.

With longer forecasts, this volatility increased, and its performance was less stable, especially after more than two hours of forecasting.

The figure shows a comparison between actual glucose values and the predicted values using the AdaBoost model. We notice that the model attempts to follow the general trend of changes in glucose levels, but sharp fluctuations are not captured accurately enough, suggesting that the model may be limited in handling complex temporal data compared to more specialized models like LSTM.

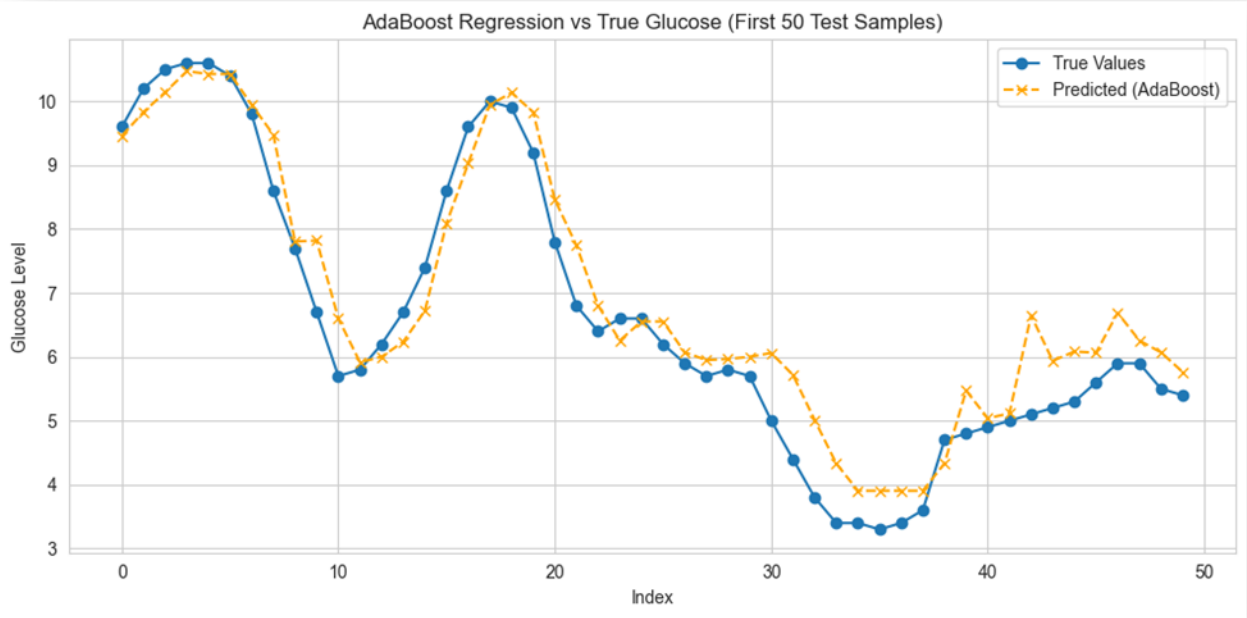

Figure 10: prediction AdaBoost regression vs true glucose.

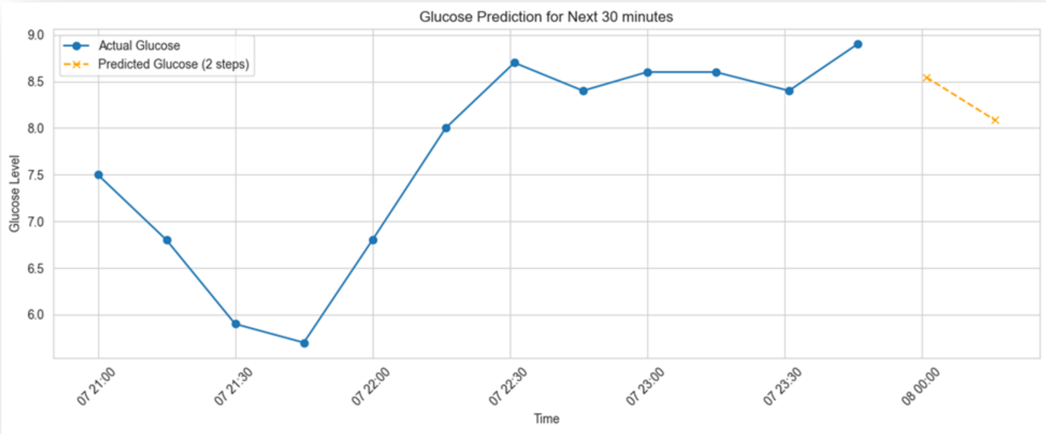

Figure 11: prediction AdaBoost for 30 minutes.

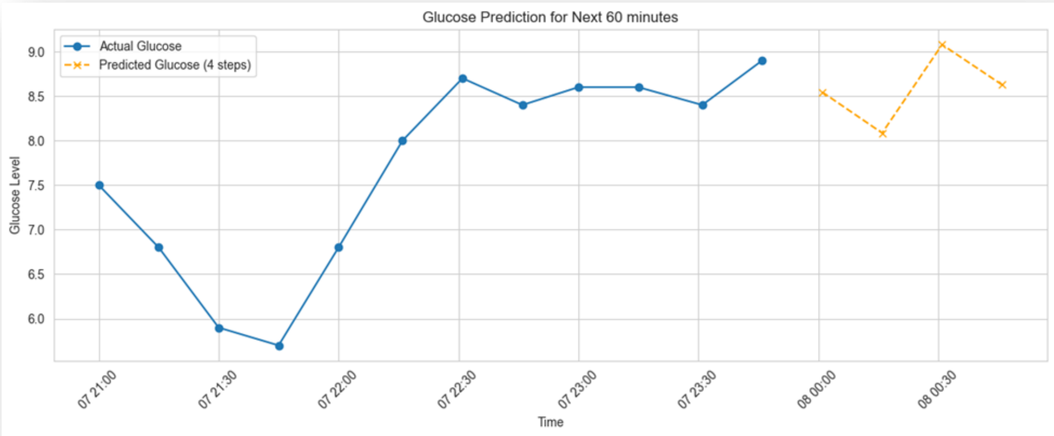

Figure 12: prediction AdaBoost for 60 minutes.

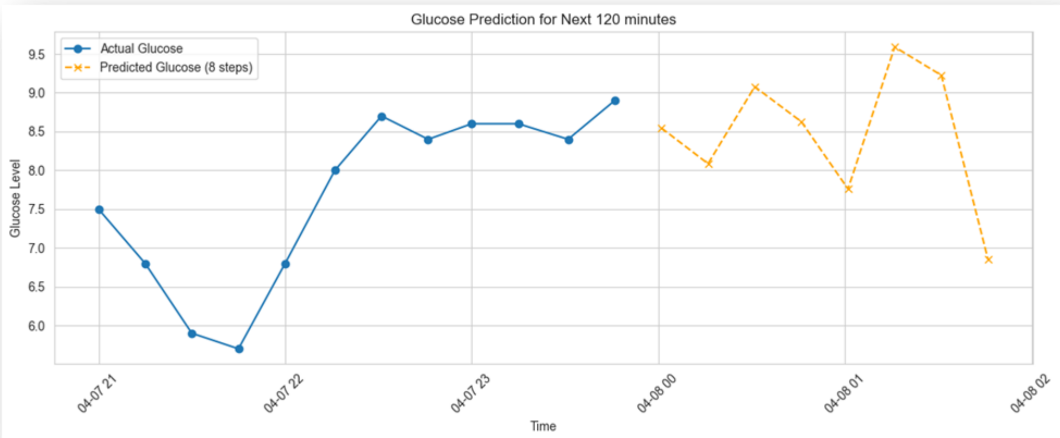

Figure 13: prediction AdaBoost for 120 minutes.

- **Forecasting beyond two hours**

In general, traditional models demonstrated acceptable forecasting ability for short periods (up to 60 minutes), while their performance gradually deteriorated as the forecast period increased, especially for forecasts beyond two hours.

This decline in accuracy underscores the need for more advanced models capable of understanding complex patterns in biometric data, such as neural network models that will be analyzed later in the project.

In forecasts beyond two hours, significant deviations between predicted and actual values were evident, highlighting the limitations of this model in handling complex temporal patterns.

In forecasts over two hours, significant deviations appeared.

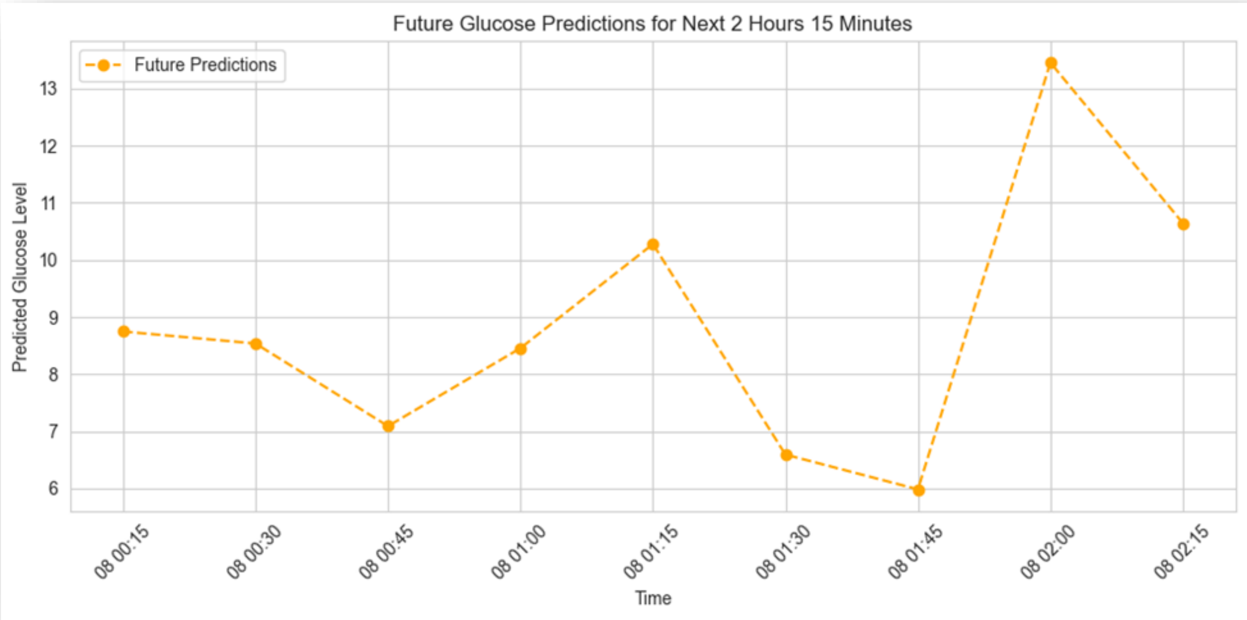

Figure 14: prediction for next 2h and 15 minutes.

7.2 Proposing new models

- **Gated Recurrent Unit (GRU) Model:**

The GRU model, short for Gated Recurrent Unit, is a type of recurrent neural network (RNN) designed to process sequential data, such as glucose data that changes over time. This model was proposed to compare its performance with that of the LSTM model, as it retains the ability to learn from long-term temporal relationships but uses fewer parameters, making it faster to train and more efficient in terms of performance.

The GRU model was adopted as an additional option alongside previous models (such as LSTM and FFNN) to:

Improve prediction accuracy by leveraging its ability to track the sequence of previous readings.

Reduced computational complexity compared to LSTM while maintaining learning quality.

Provided an alternative model that could be compared to other models in terms of performance, helping to select the most appropriate model for real-life applications.

The inclusion of the GRU model allowed the project to expand its evaluation and experimental horizons, and the model demonstrated good ability to understand temporal patterns in glucose levels, especially with data accurately resegmented into 51-minute time intervals.

- **Analysis of GRU Model Results:**

In an effort to improve the accuracy of glucose level prediction, advanced neural network models were relied upon, specifically the Gated Recurrent Unit (GRU), due to its ability to efficiently handle time series and retain long-term information.

- **Data Preparation:**

The same set of temporal features used with traditional models was prepared, with the data transformed into the appropriate format for 3D neural models (samples × time steps × features).

Each model was trained on sequential data to predict glucose levels after different time intervals: 30 minutes, 60 minutes, and 120 minutes.

It delivered performance very close to that of LSTM, with some improvement in training speed and reduced complexity.

Over short periods, predictions were accurate and stable. Although performance deteriorated slightly over longer periods, the GRU proved effective at handling biological data and adapting quickly to changes.

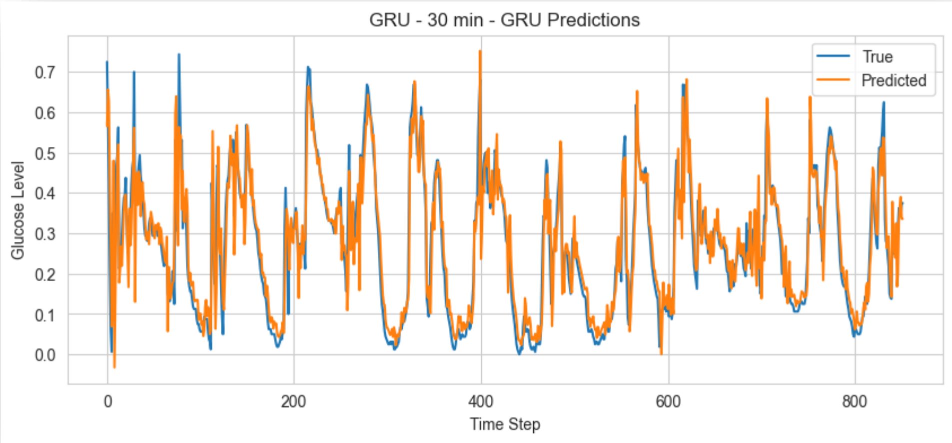

Figure 15: prediction GRU for 30 minutes.

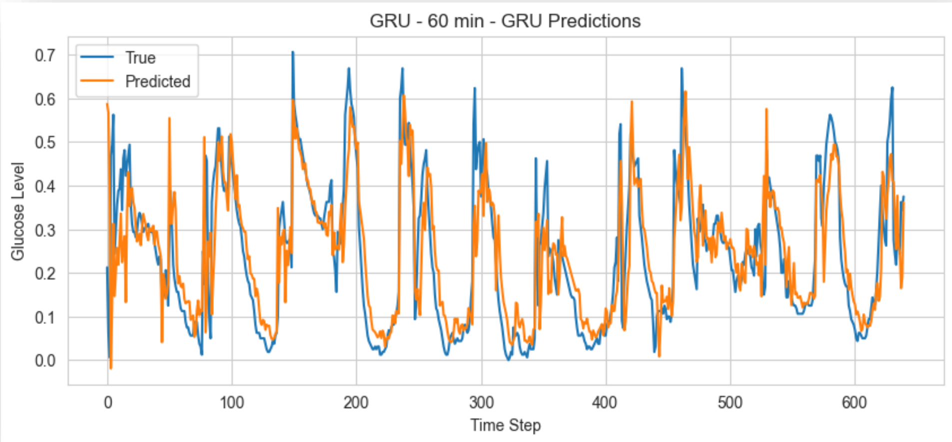

Figure 16: prediction GRU for 60 minutes.

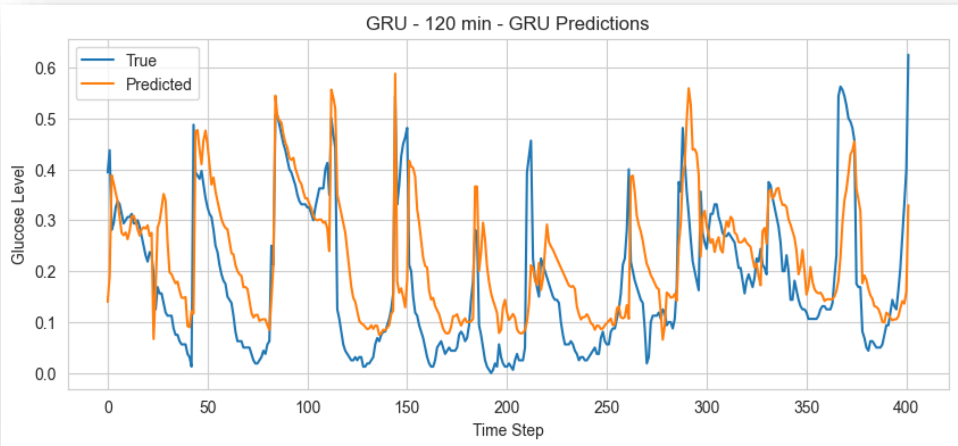

Figure 17: prediction GRU for 120 minutes.

8 Result and Comparison:

The LSTM model outperformed GRU, although the difference was not significant. However, the RMSE results confirmed this.

Neural network models clearly outperformed traditional regression models, especially over long and complex time periods.

Their ability to recognize subtle patterns and retain historical information helped them provide more accurate and realistic predictions, making them a preferred choice in intelligent systems for monitoring diabetes patients.

Model	RMSE (30 m)	RMSE (60 m)	RMSE (120 m)
GRU	0.0620	0.0942	0.1144
LSTM	0.0566	0.0712	0.0841

Table 3: Comparing results for each time series model.

9 Simple interface

One of the most important additions I made to the project was designing and building an interactive user interface using the Streamlit library. This interface aims to facilitate interaction between the user and the model. The user does not have to write programming commands. Instead, they can simply upload a data file, adjust some options, obtain results, and then generate predictions of blood glucose levels over future time periods based on the previously trained deep learning model. This interface aims to make the model easy to use by individuals not specialized in programming, such as patients or medical staff, in a simple and seamless manner.

Recommendations are generated immediately.

The interface begins by requesting the upload of a CSV file containing data over a regular time period (every 15 minutes), including key columns: glucose level, effective insulin dose (IOB), and carbohydrate intake.

Once the file is uploaded:

The model generates predictions for future glucose levels, and the results are displayed numerically and graphically for ease of understanding. Additionally, I added a feature to convert results from mmol/L to mg/dL to suit all users, depending on their system.

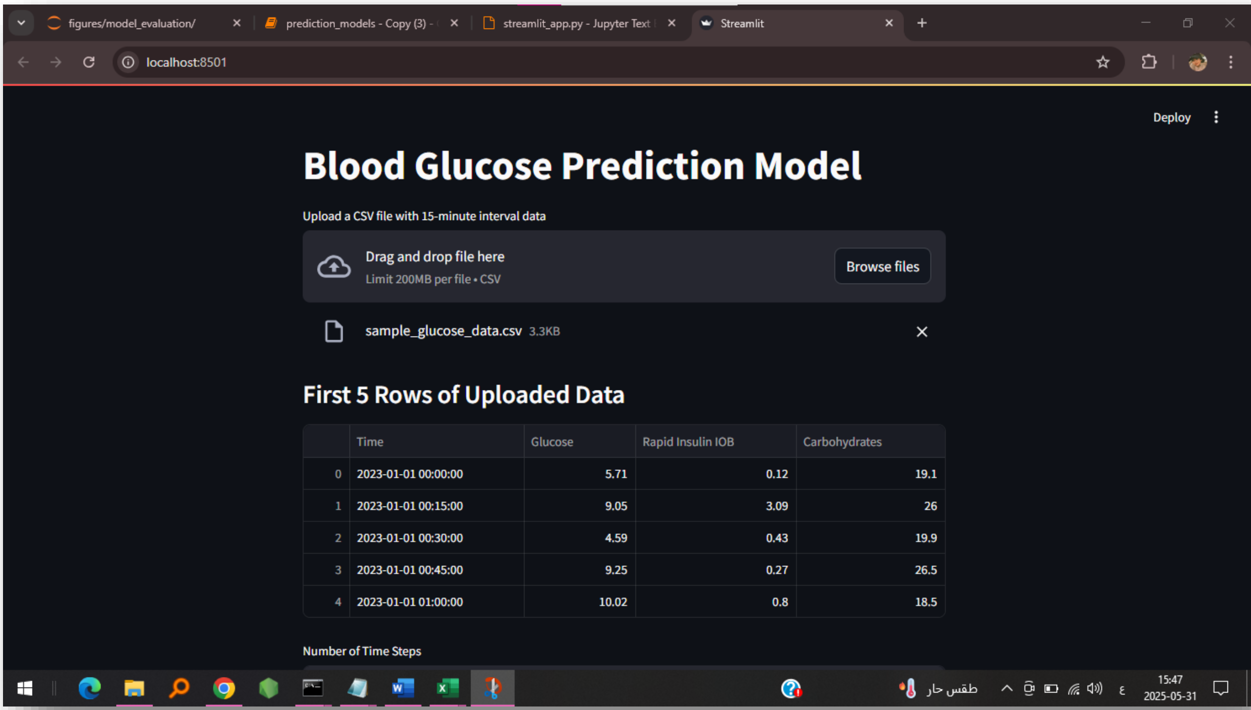

Figure 18: Interactive User Interface.

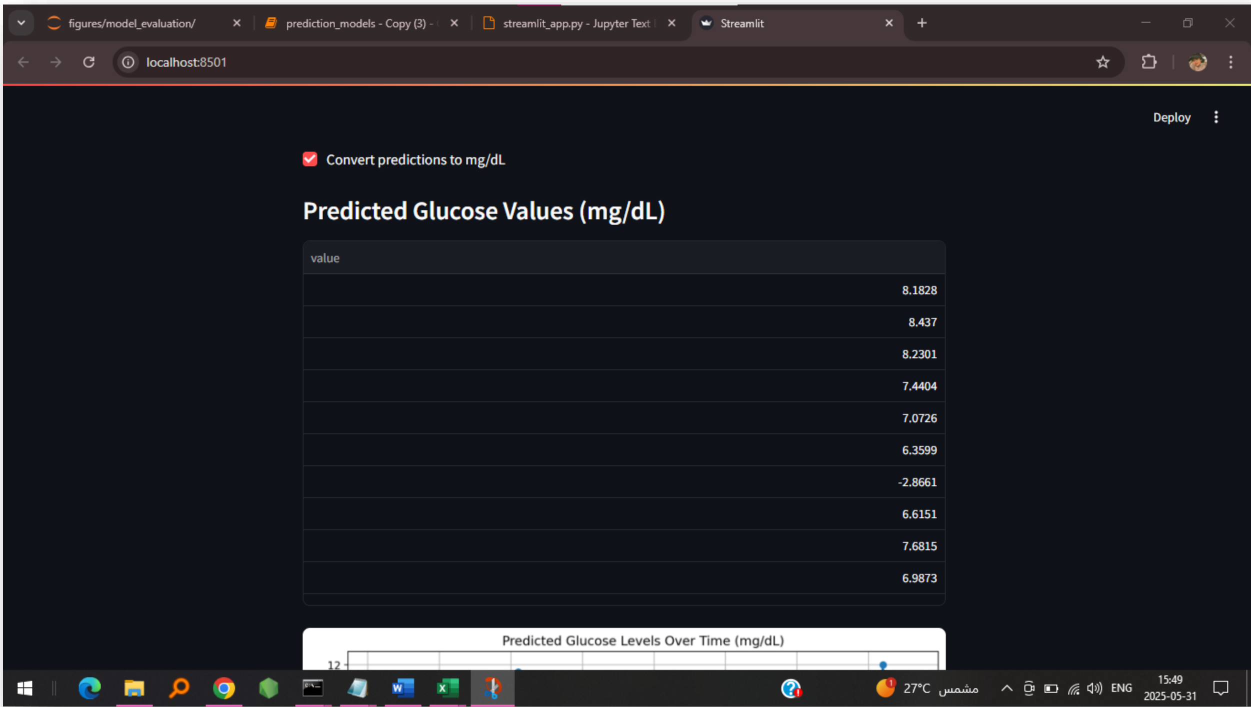

Figure 19: convert results from mmol/L to mg/dL.

Figure 20: Example of Predictions.

10 Recommendation

It doesn't stop there. After the results are displayed, automatic warning indicators appear if the predicted values indicate hypoglycemia or hyperglycemia. A dedicated button titled "📌 Show recommendations based on results" has been added, which provides the user with intelligent treatment recommendations based on the predicted condition. These recommendations are enhanced with emojis to attract attention and facilitate understanding, such as:

⚠ Danger alert.

✅ Treatment recommendation.

🎯 Indicates that the condition is stable. If all predicted values are within the normal range (70 - 180 mg/dL).

The code has been designed to allow future expansion of recommendations to include additional variables such as time of day, physical activity, or even psychological state, with the goal of providing a personalized and accurate experience for each user.

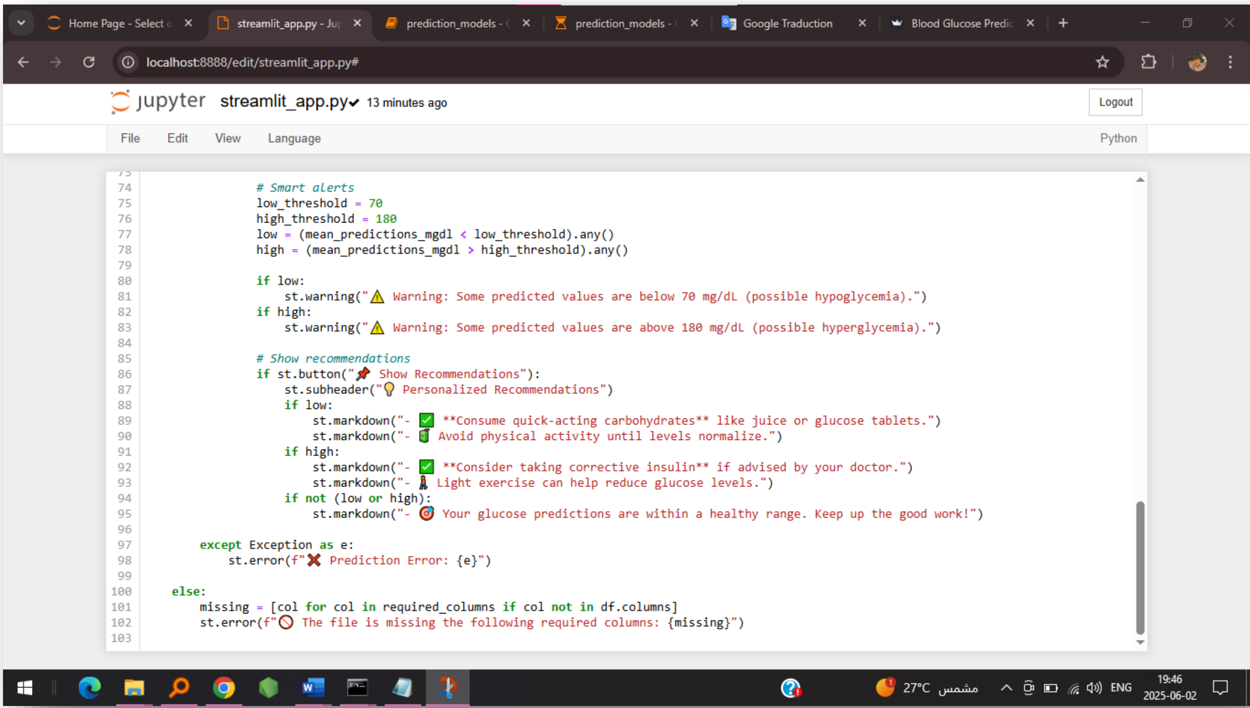

Figure 21: Streamlit interface page code.

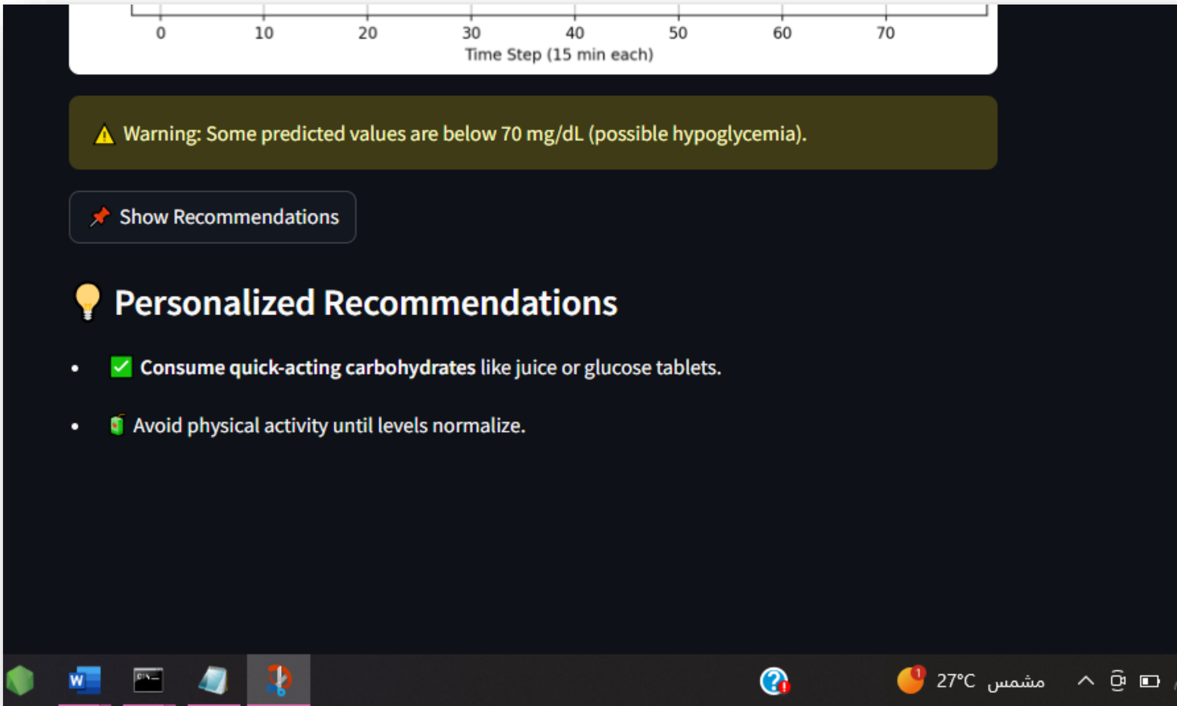

Figure 22: Example of Recommendations.

It can be said that these improvements have added real value to the project, both in terms of the accuracy of the results and in terms of ease of use for the end user, which enhances the realism of the application and its ability to be used in daily life to intelligently monitor diabetes cases.

11 Perspectives

We hope for important future developments to the project, including:

The ability to enter custom scenarios by the user, such as current glucose levels, intended carbohydrate intake, insulin dosage, and the precise timing of the event. Based on this information, the code automatically generates a data row that mimics the training data format, which is then sent to the model that performs the prediction. This opens the door to providing real-time personalized predictions that help the user make data-driven decisions.

We are using a 5-minute interval and testing its effect, as many continuous glucose monitors (CGMs) record data at intervals close to 5 minutes. This resegmentation makes the data more homogeneous and easier to process and model. The short interval reduces long time gaps between readings, which helps reduce scatter or noise in the data and improves the model's learning ability. In short, segregating the data into 5 minutes makes the model more sensitive to changes, more consistent with real-world situations, and enhances prediction accuracy and decision-making.

Generalizing the model to include data from multiple patients, as their responses to food, insulin, and physical activity vary, in addition to the type of diabetes and the number of records available for each patient. To expand the model and make it general, the patient identifier (patient_id) can be introduced into the data, and strategies such as personalized learning, adaptive transfer, and similarity clustering can be adopted. It is also important to standardize the temporal format and normalize the data to each patient. Studies recommend using multi-patient datasets such as OhioT1DM, which provide rich and diverse data, enabling the training of a comprehensive model with the ability to be customized later. Developing the current model to include multiple datasets is a practical step toward generalizing it and improving its accuracy across various clinical scenarios.

12 Conclusion

In this module, we embarked on a comprehensive analytical and practical journey that began with identifying the real problem of the difficulty of predicting glucose levels in diabetic patients, followed by a deeper understanding of the nature and complexities of the disease. We then studied the original project, which served as the basis for our development.

The various steps I took to improve the project were reviewed, starting with exploring the data and its sources, processing and cleaning it, and then applying traditional regression models and advanced neural network models such as GRU. Comparisons between the models demonstrated the importance of each, depending on the type of data and the desired prediction horizon. The analysis was further supported by an interactive interface that assists the end user, along with providing intelligent recommendations based on the predicted results.

Finally, we explored some of the future prospects for developing this work, enhancing its effectiveness and increasing its medical and technical value. This module does not represent the end goal; rather, it serves as a cornerstone for continuing research and development in the field of intelligent glucose prediction and supporting diabetic patients in a more accurate and interactive manner.

Bibliography

[1] Statista, "Artificial Intelligence - Worldwide," Statista, [Online]. Available: https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide

[2] Hsoub Academy, "Introduction to Artificial Intelligence," Hsoub Academy, [Online]. Available: https://academy.hsoub.com/programming/artificial-intelligence/%D8%A7%D9%84%D8%B0%D9%83%D8%A7%D8%A1-%D8%A7%D9%84%D8%A7%D8%B5%D8%B7%D9%86%D8%A7%D8%B9%D9%8A/.

[3][2]

[4][3] Michael Haenlein1 and Andreas Kaplan “Book A Brief History of Artificial Intelligence” https://bpb-us-w2.wpmucdn.com/sites.uab.edu/dist/6/536/files/2024/08/A-brief-history-of-AI.pdf

[5][4] Also known as: history of AI Written by B.J. Copeland Fact-checked by The Editors of Encyclopædia Britannica Last Updated: Feb 25, 2025 • Article History https://www.britannica.com/science/history-of-artificial-intelligence

[6][5] https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible

[7][6] By Hoda Jabbour January 30, 2023 https://academy.hsoub.com/programming/artificial-intelligence/%D9%85%D8%AC%D8%A7%D9%84%D8%A7%D8%AA-%D8%A7%D9%84%D8%B0%D9%83%D8%A7%D8%A1-%D8%A7%D9%84%D8%A7%D8%B5%D8%B7%D9%86%D8%A7%D8%B9%D9%8A/

[8][7] https://www.argaam.com/ar/article/articledetail/id/1615530

[9][8] https://www.jpl.nasa.gov/news/heres-how-ai-is-changing-nasas-mars-rover-science/

[10][9] https://www.theinformation.com/articles/google-plans-ad-sales-restructuring-as-automation-booms?

[11][10] By Ahmed Essam El Naggar March 19 2023 https://academy.hsoub.com/programming/artificial-intelligence/%D9%86%D8%B8%D8%B1%D9%8A%D8%A9-%D8%B3%D8%B1%D9%8A%D8%B9%D8%A9-%D8%B9%D9%84%D9%89-%D9%85%D8%AC%D8%A7%D9%84-%D8%AA%D8%B9%D9%84%D9%85-%D8%A7%D9%84%D8%A2%D9%84%D8%A9-r1933/

[12][11] https://diabetes.org/about-diabetes/type-1

[13][12] https://www.who.int/health-topics/diabetes#tab=tab_1

[14][13] https://diabetesjournals.org/care/article/44/Supplement_1/S15/30859/2-Classification-and-Diagnosis-of-Diabetes

[15][14] https://diabetes.org/about-diabetes/type-1

[16][15] https://github.com/MrBlueHere/Blood-Glucose-Prediction/blob/main/thesis_text/thesis.pdf

[17][16] FOSTER-POWELL, K; MILLER, J B. International tables of glycemic index. The Ameri

can Journal of Clinical Nutrition. 1995, vol. 62, no. 4, 871S–890S. issn 0002-9165. Available

from doi:
https://www.sciencedirect.com/science/article/abs/pii/S0002916523190289?via%3Dihub

~~[18]~~[17] DAVIDSON, Paul C; HEBBLEWHITE, Harry R; STEED, Robert D; BODE, Bruce W.

Analysis of guidelines for basal-bolus insulin dosing: basal insulin, correction factor, and carbohydrate-to-insulin ratio. Endocrine Practice. 2008, vol. 14, no. 9, pp. 1095–1101. Available also from: https://doi.org/10.4158/EP.14.9.1095.

~~[19]~~[18] https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001189.v1.p1

~~[20]~~[19] https://ieee-dataport.org/

~~[21]~~[20] https://www.kaggle.com/search?q=blood+glucose+prediction

~~[22]~~[21] H. Lee, H. Lee, and S. Yoon, "deepCGM: Convolutional Neural Networks for Predicting Continuous Glucose Monitoring Data," Sensors, vol. 20, no. 17, pp. 1–13, Sep. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/17/4817

~~[23]~~[22] A. Mirshekarian, Q. Li, J. You, and Z. Yu, "Diaredict: An interpretable deep learning framework for diabetes management," IEEE J. Biomed. Health Inform., vol. 26, no. 2, pp. 535–545, Feb. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9583737

~~[24]~~[23] X. Zhou, Z. Yu, and Z. Wang, "GlucoNet: A Hybrid Deep Learning Model for Blood Glucose Prediction," in Proc. Int. Joint Conf. Artif. Intell. (IJCAI), Montreal, Canada, 2021. [Online]. Available: https://www.ijcai.org/proceedings/2021/0497.pdf

~~[25]~~[24] D. Choi and L. Floriš, "Predicting Blood Glucose Levels with LMU-RNNs," GitHub Repository, 2023. [Online]. Available: https://github.com/dongsikchoi/Blood-Glucose-Prediction-LSTM

~~[26]~~[25] MrBlueHere, "Blood Glucose Prediction Using LSTM," GitHub Repository, 2023. [Online]. Available: https://github.com/MrBlueHere/Blood-Glucose-Prediction

~~[27]~~[26] Glucose Tolerance test [online] [visited on 2022-04-18]. Available from: https://www.medlineplus.gov/ency/article/003466.htm.

~~[28]~~[27] CRYER, Philip E.; DAVIS, Stephen N.; SHAMOON, Harry. Hypoglycemia in Diabetes.

Diabetes Care. 2003, vol. 26, no. 6, pp. 1902–1912. issn 0149-5992. Available from doi:

https://diabetesjournals.org/care/article/26/6/1902/26493/Hypoglycemia-in-Diabetes

~~[29]~~[28] Hyperglycemia [online] [visited on 2022-04-18]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK430900/

~~[30]~~[29] KROLL, Martin H. Biological variation of glucose and insulin includes a deterministic

chaotic component. Biosystems. 1999, vol. 50, no. 3, pp. 189–201. issn 0303-2647.

Available from doi:
https://www.sciencedirect.com/science/article/pii/S0303264799000076?via%3Dihub

~~[31]~~[30] DINE4FIT, A.S. Kalorick´e Tabulky [comp. software]. 2022. Version 3.7.1. Available also
from: https://www.kaloricketabulky.cz/

الملخص :

في ختام هذا المشروع، تتجلى أهمية الذكاء الاصطناعي وتحديدًا تقنيات التعلم الآلي والتعلم العميق في تقديم حلول واقعية وفعالة لمشكلة طبية مزمنة كالسكري. لم يكن الهدف مجرد تطوير نموذج لتوقع مستويات الجلوكوز، بل كان طموحنا أن نقدم أداة تساعد المرضى في اتخاذ قرارات علاجية مدروسة، وتخفف من عبء المتابعة اليومية الثقيلة. من خلال الجمع بين التحليل العميق للبيانات، واختبار عدة نماذج، وتصميم واجهة تفاعلية توصل النتائج ببساطة، أضفنا بُعدًا إنسانيًا لتطبيق علمي. ورغم أن هذا العمل ليس نهاية الطريق، إلا أنه يشكل خطوة واثقة نحو مستقبل تُسخر فيه خوارزميات تعلم الالة والتعلم العميق لتحسين جودة حياة المصابين بمرض السكري عبر أنظمة تنبؤ ذكية وموثوقة.

الكلمات المفتاحية: توقع مستويات الجلوكوز في الدم، السكري، التعلم الآلي، التعلم العميق

~~Conclusion~~Abstract:

As we reach the end of this journey, the value of artificial intelligence especially Machine Learning and Deep Learning in addressing chronic medical challenges becomes unmistakably clear. This project was never just about predicting blood glucose levels; it was about empowering diabetic patients with a smarter, more personalized way to manage their condition. By blending scientific rigor with practical implementation and user centric design, we’ve taken a meaningful step toward a future where data driven insights support daily healthcare decisions. While there is still much to explore, this work sets a solid foundation for further innovations in blood glucose prediction and intelligent medical assistance systems.

Keywords: Blood glucose prediction, Diabetes, Machine Learning, Deep Learning

~~conclusion~~Résumé:

Au terme de ce projet, il devient évident que les technologies d’intelligence artificielle, notamment le Machine Learning et le Deep Learning, offrent des perspectives concrètes pour relever les défis liés au diabète. Notre objectif n'était pas simplement de prédire les niveaux de glycémie, mais bien de fournir aux patients un outil intuitif capable de les accompagner dans la gestion quotidienne de leur maladie. En combinant une analyse rigoureuse des données, l’évaluation de plusieurs modèles et une interface conviviale, nous avons posé les bases d’une solution humaine et efficace. Ce projet n’est pas une finalité, mais une première étape vers des systèmes intelligents capables d'améliorer la vie des personnes atteintes de diabète grâce à la prédiction du glucose sanguin.

Mots-clés : Prédiction du glucose sanguin, Diabète, Machine Learning, Deep Learning