

Toward an Interpretable Continuous Glucose Monitoring Data Modeling

Juan Francisco Gaitán-Guerrero¹, José Luis López Ruiz¹, Macarena Espinilla¹,
and Carmen Martínez-Cruz¹

Abstract—The ongoing global health challenge posed by diabetes necessitates a critical understanding of all generated data streamed from sensors. To address this, our study presents a robust fuzzy-logic-based descriptive analysis of glucose sensor data. This analysis is embedded within the context of an innovative architecture designed to support multipatient monitoring, with the goal of assisting healthcare professionals in their daily tasks and providing essential decision-making tools. Our novel approach captures and interprets complex data patterns from glucose sensors, and also introduces the capability of creating high-quality linguistic summaries, to highlight the most relevant phenomena through the use of natural language (NL). These descriptions facilitate clear communication between healthcare professionals and people with diabetes, enhancing a deeper understanding of intricate data patterns and promoting collaboration in diabetes care. A comparative evaluation between our proposal and the one obtained using GPT-4 underscores the sustainability, effectiveness, and efficiency of our methodology, positioning it as a new standard for empowering diabetic patients in terms of care and prevention, contributing to their progress and well-being.

Index Terms—Diabetes, fuzzy logic (FL), GPT-4, GPT-4o, Internet of Medical Things (IoMT), Internet of Things (IoT), linguistic descriptions of time series (TS), linguistic summaries, medical devices, natural language (NL) generation.

I. INTRODUCTION

THE GLOBAL surge in diabetes worldwide, in an age where health concerns deeply, has given rise to the proposal of different strategies in order to deal with this quite extended disease. Regarding diabetes, a medical viewpoint defines it as a sequence of metabolic disorders linked to hyperglycemia and caused by partial or total insulin insufficiency [1], deriving in multiple long-term affections such as heart disease, nerve damage, vision impairment and urinary tract infections. It is materialized in the situation where either the pancreas is not producing insulin or when the body is no longer able to manage it; the nonglucose supply to body cells derives not only in no energy administration but also

in hyperglycemia episodes. According to the International Diabetes Federation [2], there are over 537 million people suffering from diabetes worldwide, highlighting that 3 in 4 adults are affected by it in middle-low-income countries, plus the 6.7 million deaths that were registered in 2021.

Currently, healthcare systems struggle to provide personalized care due to a number of challenges [3], including a technological limitation. This type of care requires advanced technologies such as continuous glucose monitoring. In the case of patients with diabetes, it happens that many data is generated through the use of wearable glucose sensors. Consequently, a big effort from the healthcare professionals is a must in an attempt to describe the extensive retrieved data; it is a challenge to efficiently manage and interpret the dynamic flows of glucose levels. Hence, there arises the need to supply both patients and clinicians with tools that provide personalized care.

The proliferation of enormous amount of information generated through Internet of Things (IoT) devices has led to the need of synthesize the collected raw data. An IoT system materializes by the time an interconnected network of physical objects or entities is established, enabling the collection of specific physical phenomena data through the utilization of one or multiple sensors connected to the Internet [4]. These are the premises that have paved the way for an impressive and diverse flow of information in the actual society, where the medical field is witnessing a massive production of electronic data [5], as characterized by a highly intricate analytical process. Nowadays, IoT technology is applied in a wide range of applications: in smart cities [6], [7], in smart grids [8], in precision agriculture [9], in smart vehicles [10], and in healthcare [11].

In the healthcare field, different types of sensors are playing a crucial role in safeguarding human lives, under the paradigm of the Internet of Medical Things (IoMT), which refers to the collection of medical devices that are interconnected within a healthcare network, streaming high-risk and valuable data [12]. Particularly in the case of diabetic patients, in the last decade there has been a proliferation of sensors to measure glucose in real time, facilitating glucose monitoring. However, these sensors are not integrated into the healthcare system [13], [14], [15], [16]. Thus, it is difficult for healthcare professionals to access and evaluate patient data and provide personalized care.

On the other hand, such systems generate over time a large amount of data that must be processed. In order to manage all the information generated, traditionally used tools

Manuscript received 4 March 2024; revised 27 May 2024; accepted 11 June 2024. Date of publication 26 June 2024; date of current version 25 September 2024. This result has been partially supported by Grant PID2021-127275OB-I00 and Grant PID2021-126363NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by “ERDF A way of making Europe.” Funding for open access charge: Universidad de Jaén. (Corresponding author: Juan Francisco Gaitán-Guerrero.)

The authors are with the Department of Computer Science, University of Jaén, 23071 Jaén, Spain (e-mail: jgaitan@ujaen.es; llopez@ujaen.es; mestevez@ujaen.es; cmcruz@ujaen.es).

Digital Object Identifier 10.1109/JIOT.2024.3419260

do commonly refer to those sentencing the validity or not of a statement. Logic only accepts values in the domain true or false. In set theory, a value belongs to a set or not. An optimization problem has an answer or not [17]. According to American Diabetes Association (ADA) [18], different thresholds are defined aiming to diagnose the disease. Despite that, glucose levels are commonly communicated in a qualitative manner, resulting in different expert-based-statements influenced by their own point of view when labeling the value. This gives rise to assume the uncertainty and lack of clarity that characterize glucose-monitoring-collected-data over time. Considering the scenario where only one sample has been registered, the following example can be used to explain this easily: a patient is attending a doctor appointment and doctor states “Your glucose level is too high, you have developed diabetes,” the person would immediately assume these words, both ignoring different factors that may have influenced the detected value, i.e., a stress situation or a recently done physical effort. In this context, fuzzy logic (FL) has a very important role to play in modeling this uncertainty.

In fact, FL theory [19] has allowed to represent data and compute on them in a flexible way (see compilations [20]), with the following benefits in this context.

- 1) Minimizing the impact of data loss or corruption due to technical problems associated with IoT devices.
- 2) Blurring the validity ranges of a measure, so as to resemble how humans communicate and understand the world, as shown in the previous example.
- 3) Summarizing large amounts of data in a semantic approach closer to human perception.

Contrary to what is observed in the literature, where in the domain of diabetes, various and numerous proposals have encompassed the binary classification of diabetic patients or the prediction of future values. Liu et al. [21] proposed a machine learning method for a future 15–60 min glucose values prediction considering a 30-min previous sampling, and determining accelerometry as an important factor to improve the predicted values. Rastogi and Bansal [22] contributed with the study of different data mining techniques aiming to predict the disease while pretending to preserve interpretability. In [23], a novel neural architecture for glucose and meal prediction is evaluated using a probabilistic approach. It is notable how machine learning methods present important advantages in glucose and diabetes prediction, as often characterized by interpretability and the need for a smaller number of samples for learning. Nonetheless, pattern identification may be difficult when data sets are large and complex. Therefore, deep-learning techniques provide assistance to these kinds of situations, while interpretability is left unexploited. Despite all, mechanisms for diabetes diagnosis and glycemia prediction continue to be successfully implemented. However, it is acknowledged that a more nuanced comprehension of the patient’s glucose variations is essential for disease diagnosis and treatment determination.

In this regard, an innovative approach that generates linguistic descriptions of time series (TS) from glucose sensors, is also presented in this contribution; it is provided a set of relevant information from an interpretable model to the

healthcare professionals that allows them to perceive how the statements have been generated without a significant cognitive effort. For this purpose, an IoT-based architecture is defined for the acquisition of glucose data in real time, independently of the number of patients, as opposed to other proposals. Finally, when assessing the developed system, it is found as an imperative to analyze end-users’ perception of the generated outputs through the utilization of surveys, serving as a pivotal strategy for evaluating the overall performance. This evaluation is complemented with a comparison of our proposal and the output obtained from large language models (LLMs), such as GPT-4 [24], [25]. Consequently, we highlight the main contributions of our work as follows.

- 1) *Multipatient real-time monitoring architecture* designed for data collection at any time, place, and for anyone, ensuring data storage.
- 2) *Linguistic summarization of data methodology* unprecedented in the literature, utilizing a rule-based system for semantically communicate relevant expert-guided statements.
- 3) *Integrated framework for data and natural language (NL) summaries visualization* combining the proposed architecture and the linguistic summarization methodology for comprehensive and efficient data interpretation and healthcare assistance.

The document is divided into the following sections. First, a review of related works can be found in Section II. Next, Section III defines the IoT-based monitoring system for glucose data acquisition. Afterward, the FL-based approach to generate linguistic summaries is proposed in Section IV. Subsequently, the evaluation of the approach is performed in Section V, and finally the conclusions and future work are discussed in Section VI.

II. RELATED WORKS

In the context of glucose monitoring in diabetic patients, the literature provides a large number of studies related to this type of system [15]. These works employ IoT devices in a multitude of scenarios in which not only biometric data are used. Many of these proposals include the use of wristbands, smartwatches, and other types of wearables that provide information about the health status of diabetic patients [26]. Other studies also establish the use of smartphones for detection purposes [27]. Thus, automatic activity recognition devices are the most widely used, while glucose monitoring sensors are following, highlighting the following proposed systems [13], [14], [16].

In contrast to our proposal, these works focus on defining a methodology, but do not establish a real architecture for gathering glucose data or omit some details, such as the sensor used. Furthermore, neither do they indicate whether the system can be used at any site nor if the proposal is multipatient.

For their part, when describing monitored data, linguistic uncertainties prompt consideration of the utilization of FL, as it extends classical set theory in a way that registered values do belong to a set according to a membership degree, illustrated

in the fuzzy sets (FS) theory proposed by Zadeh [19]. This FS capability emerges as a powerful tool for capturing these lacks of precision inherent in oral communication. Zadeh postulated computing with words (CW) and computational theory of perceptions (CTP) [28] theories seeking to bridge the gap between imprecise, vague human NL and the precise, binary nature found in computation. Yager's proposal of linguistic summarization of data through the use of fuzzy quantified sentences was influenced by these paradigms [29], displaying the knowledge inherent in databases. Yager [30] also introduced indispensable elements taking part in linguistic summarization including these three.

- 1) *Summarizer (S)* is a label or linguistic value that is defined in a domain of an attribute within a data set D (e.g., high).
- 2) *Quantity in agreement (Q)* to indicate the amount of data satisfying S (e.g., many).
- 3) *A measure of truth of the summary [Degree of Truth (DoT)]* to establish the truth of the statement involving Q individuals satisfying S in the domain D .

Zadeh [19], [31] introduced this way, the use of protoforms (prototypical forms) whose application has been dedicated to the knowledge extraction from TS. Two types are considered differing as follows.

- 1) *Protoform 1* structured in the form A is S where A is an attribute in the data set D and S is the summarizer (e.g., glucose level is high).
- 2) *Protoform 2* comes to quantify through Q a type 1 protoform, leading to a Q A are S structure (e.g., many glucose levels are high).

Addressing the latter, some relevant contributions in the medical field are mentioned: Mehra et al. [32] introduced an IF-THEN rule generator system for the description of a medical diabetes database with the utilization of NL. Along the same line, Amghar and Chikh [5] studied the way of generating a linguistic summary from a data set, serving as the input of a classification model that could determine the state of a patient (diabetic or nondiabetic) thanks to fuzzy cardinality, obtaining this way a fully interpretable model at last. These experiments, capable of doing classification, do actually consider samples that are registered at a specific point, not contemplating a fully fledged collection of glucose levels. Nasser et al. [33] have contributed with a deep-learning algorithm capable of predicting glucose levels in patients that have been monitored. To do so, 20 previous samples are considered to prevent harmful hyperglycemia episodes in a future period of 30 min. Once again, the number of samples taken into account are not reliable enough as the regard of external factors (e.g., physical effort, food ingestion, or insulin supply) is an imperative.

Focusing on the generation of linguistic descriptions of TS (GLiDTS), many authors have successfully devised different systems for data condensation which provide a nonredundant analysis of collected TS; data abstraction is then applied facilitating comprehensive NL summaries across diverse fields. Marín and Sánchez [34] enriched the literature considering in their proposal the GLiDTS as a technique, different from other tasks dedicated to realizing TS analysis, defining it

as systems able to simulate the answer of a human expert when asked about what he/she can observe or highlight in the TS. For this process, the authors described it by the analysis of TS for feature extraction, mostly represented by protoforms and later, by generating a summary using NL. The use of machine learning methods is contemplating when addressing the second task, however, their correlation with certain scenarios is highlighted, specifically, where it is required the identification of association rules, fuzzy approximate or gradual dependencies, and/or anomalies in rules. Despite this, there are application domains where such solutions are not suitable, as the creation of predefined protoforms becomes necessary, aiming to meet the requirements of the system. Recent solutions of this kind are cited next: Kaya et al. [35] have contributed with a linguistic summarization approach of EEG signals for emotional responses to advertisements description. Cascallar-Fuentes et al. [36] presented a fuzzy temporal ontology to facilitate the interpretation of TS, applying it to air quality index data, and providing comprehensive linguistic descriptions at last. In [37], meteorological data implicit knowledge is modeled for the generation of protoforms that create a big search space which is later explored by two different meta-heuristics. The economic field has also taken benefit from the work of Genc et al. [38] where relationships and interactions within an international trade network are linguistically summarized. Responding to the definition of GLiDTS, Hudec et al. used fuzzy modeling of statistical data to make it accessible and understandable, using linguistic summaries, for a wider audience.

Concretely, in the health realm, Peláez-Aguilera et al. [39] developed expert-guided fuzzy linguistic protoforms for summarizing heart rate streams obtained from a smart wrist-worn gadget in cardiac rehabilitation sessions. Albín-Rodríguez et al. [40] also established different fuzzy protoforms when evaluating hyperactive behavior. This proposal comprehends the aggregation of data flows streamed from wearable and mobile devices, leading to obtain a highly accurate monitoring of an individual's daily environment. Jain et al. [41] created a framework for processing home sensors data characterized by resting time, heart rate, respiration, walking speed or time in bed in elderly people; both synthetic and real data were computed in order to validate the developed fuzzy-driven NL generator (NLG). The eldercare field was also influenced by a novel approach proposed by Martínez-Cruz et al. [42]. It was effectively applied to the description of data from bed restlessness monitoring, not only providing a high semantic content but also emphasizing the key features present in TS. However, none of these prototypes specifically address real-time processing tasks. Furthermore, it is essential to highlight that the subject we propose is unprecedented in the existing literature, while holding vital importance for public health personnel; simplifying and summarizing raw data using NL is a must due to its sheer volume.

The existing literature reveals how data regarding glucose fluctuations is often dedicated for the prediction of target or future values where interpretability is, in some cases, left behind, by the use of modern tools such as deep learning [33].

This oversight highlights the necessity of providing both patients and healthcare professionals with an approach that prioritizes understanding patterns and dynamics of data during patient monitoring, rather than solely focusing on classification tasks that commonly rely on limited sampling; some machine learning methods handle data description for diabetes diagnosis not involving a complete fluid of information neither as shown in [5] and [32], which may cause a turning point in decision making.

In parallel, LLMs are found to show encouraging signs in analyzing and providing medical data as attested in [43] and [44]. As a consequence, some but few efforts have been recently put on the utilization of these tools for continuous glucose monitoring description, specially taking advantage of this technology. Healey et al. [24] underlined promising results from GPT-4, using its Data Analyst plugin, aiming to integrate this tool within diabetes care. One of the premises of this study is the proposal found in [25], where daily glucose data is pretended to be described with the chatbot. Conversely, in [24], it is desired to obtain summarizes contemplating 14-day synthetic data. A validation of the results is carried out by experts in the domain, then estimating its possible integration within glucose management systems. Despite the fact that the approach is presented with the obtaining of promising results, some mistakes are found within the 14-day summarization they provide. This underscores the necessity of conducting a daily data analysis as a prerequisite for more complex evaluations. Further discussion of this proposal is done in Section V-C.

To tackle the need to simplify a large magnitude of data flows while ensuring interpretability, this article advocates for a diabetes management framework that combines a multipatient monitoring system with a white-box rule-based model for linguistic summarization of data in real time, based on GLiDTS principals, and focusing on a daily basis.

III. MONITORING SYSTEM

In this section, the proposed IoT-based monitoring system for real-time data collection is defined. This proposal is critical to provide personalized care through linguistic summaries using our methodology. Therefore, this section first discusses the commercial glucose sensor used (Section III-A), and also the architecture of the system (Section III-B).

A. Proposed Sensor

Blood glucose measurement is very important in determining the health of a person with diabetes. This measurement refers to the blood sugar at a given time and is expressed in milligrams of sugar per deciliter (mg/dL). A glucometer is used to test blood glucose. This device allows the performance of a capillary blood glucose test using a drop of blood from a fingertip. The two instruments used are the lancet and the test strip. The lancet takes a small drop of blood from the finger and the test strip interacts with the blood to measure the glucose level. Finally, the glucometer evaluates the test strip to give the blood glucose value. Fig. 1 shows each of these components.



Fig. 1. Following tools are needed to measure blood glucose: test strip, lancet, glucometer, and needle (left to right). Example model: GlucoMen aero 2K manufactured by Menarini Diagnostics.

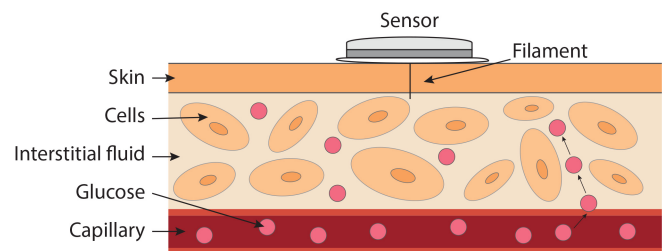


Fig. 2. Illustration of where to place a continuous glucose monitoring sensor on the human body.

Continuous glucose monitoring is now possible with commercially available sensors. These devices use glucose from interstitial fluid, which comes from the exchange between tissue cells and blood, unlike conventional glucometers that use capillary glucose.

The main difference between this and capillary glucose monitoring is that the former has a latency of 5–10 min, while the latter only occurs during steady periods. For example, when a user eats, the readings will not match and will show a small delay.

These sensors consist of a filament or electrode. It is placed subcutaneously using an applicator and works by measuring interstitial glucose levels through an enzymatic reaction. Regular readings are obtained. Fig. 2 shows an illustration of the sensor applied to human skin.

This study uses the commercial sensor Freestyle Libre 3 [45]. This commercial sensor has the following features.

- 1) Dimensions are 21 × 2.9 mm (diameter × depth).
- 2) Battery life is estimated at 14 days.
- 3) Sampling rate is 5 min.
- 4) Bluetooth low-energy (BLE) connection is used to transfer glucose data.
- 5) Nonreusable and costs \$63.26 per unit.

B. System Architecture

The next step is to define the architecture of the monitoring system. In this case, the research uses an IoT-based

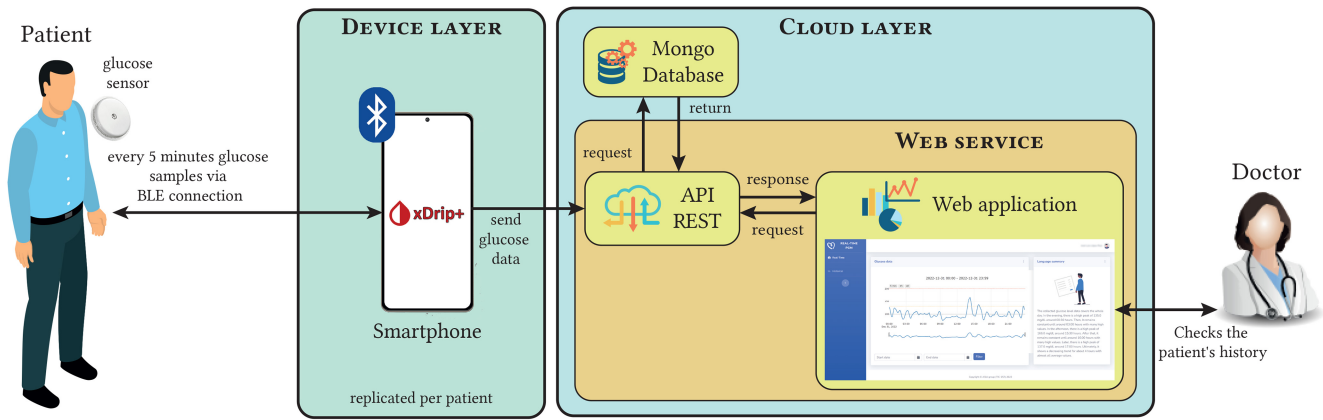


Fig. 3. IoT-based monitoring system architecture divided into fog and cloud layers.

architecture, whose purpose is to collect glucose data in real time and send it to the cloud where it is stored and processed.

This architecture can be divided in two different layers. The device layer consists of glucose sensors and mobile devices. Therefore, a continuous glucose sensor and a smartphone are set up for each patient. To connect the two devices, a pairing is first established using near-field communication (NFC) technology, and once this is done, samples are continuously sent via BLE. Therefore, the patient is not required to continuously interact with the sensor to obtain glucose data. Furthermore, the real-time factor is crucial for people with type 1 diabetes, as hypoglycemia can lead to epileptic seizures and even brain damage. The whole process is performed through a mobile application called xDrip+ [46].

On the other hand, in the cloud layer, there is a dedicated server that hosts a database, a Web application for patient monitoring that uses the methodology presented in this article to provide healthcare professionals with personalized care for their patients. The mobile device continuously sends the collected samples to the cloud through the Nightscout module which is integrated in xDrip+. For this purpose, a REST API service has been implemented that replicates the functionality of Nightscout [47]. In contrast to the original service, our system is able to receive data from different patients and recognize them based on the identifier provided.

Finally, the samples are persistently stored in the database. The monitoring Web application allows the historical data of the patient to be consulted at any time through raw data and linguistic summaries.

The whole of this architecture is shown in Fig. 3.

IV. METHODOLOGY

In this section, a novel methodology based on the computation of a linguistic summary is proposed, as characterized by the analysis of glucose-monitoring-patient TS collected in real time. The exposed methodology comes to the rescue after the different shortcomings found in the reviewed systems concerning patient monitoring and data description.

The extraction of the implicit knowledge existing in the compiled raw data, represents a nontrivial procedure of selection and modeling of information that, until further treatment,

remains unknown while being potentially useful. This process is carried out following a first task based on the knowledge discovery in databases (KDDs) procedure, with the inclusion of an NLG, as detailed in Marín and Sánchez [34] proposed architecture. Initially, data is retrieved and segmented, for a posterior characterization using summarizers, whose purpose is to describe observable features, forming protoforms that are then translated into sentences. This process, which is illustrated in Fig. 4, derives in the generation of a quality-framework-satisfied linguistic summary at last.

A. Preprocessing

First, it must be considered that different attributes are being saved within every instance registered by the glucose sensor. In the context of variable selection, our system only contemplates the inclusion of those referring to the timestamp and the glucose level collected in 5-min epochs. As a result, a TS is defined as: let $\{x_t\}_{t=1}^n$ be the sequence of observations indexed by t , $x_t : x_t \in \mathbb{R}$ represents the glucose level corresponding to time t . Data acquisition may fail at some point, so when handling gap filling, Moving Average Window algorithm is applied in intervals where $|\{x_t : x_t = \emptyset\}| < 6$, in order not to estimate periods longer than half an hour.

Graphically, a TS is represented with n-line-connected points in \mathbb{R}^2 , what derives in the idea of a geometrical top-down segmentation; partial singularities and redundancies are placed in a new homogeneous segment as a whole. Some significant data simplification techniques are found in the works of Farsi et al. [48], Ali and Aggarwal [49], or Tang et al. [50].

The segmentation method used in this work is known as Ramer–Douglas–Peucker (RDP) algorithm, as exposed in [51]. As an input, RDP receives a complete TS to be segmented, deriving in another TS with a lower number of points; steps to follow by RDP are detailed in the approach proposed by Martinez-Cruz et al. [42]. The choice of this algorithm is made because it preserves the most relevant TS features in a very simple and intuitive way. This method has been used with very satisfactory results in articles like [42] and [52]. Despite of having a large number of segmentation methods based on piecewise linear approximations, with similar or even

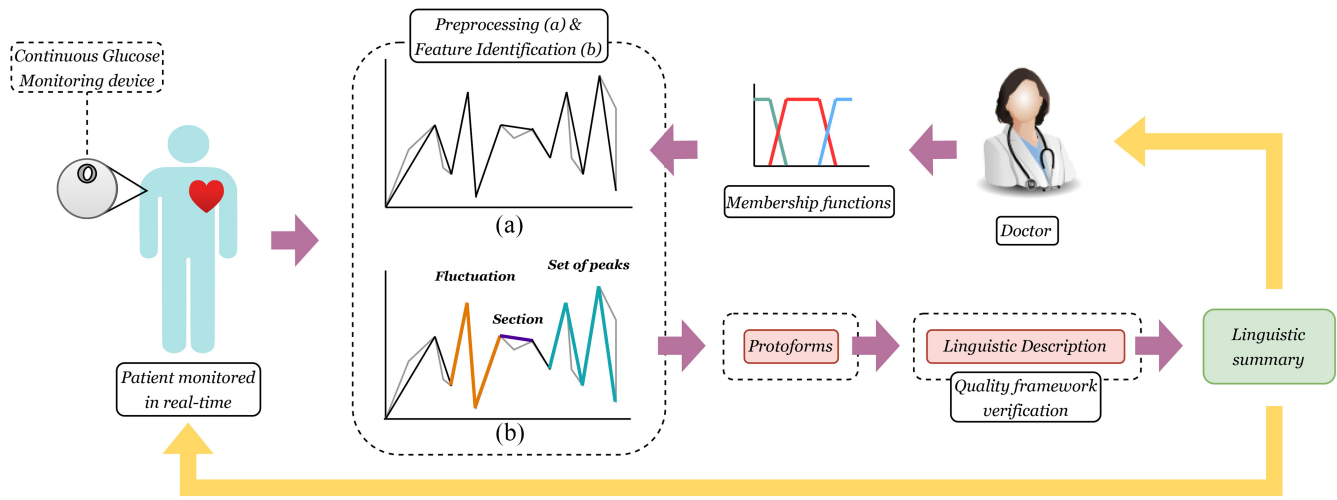


Fig. 4. Architecture of the proposed methodology.

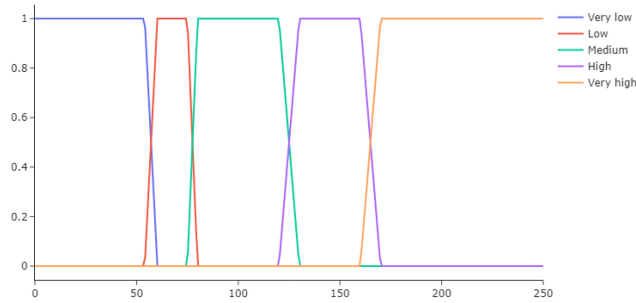


Fig. 5. Membership functions for glucose levels summarizers.

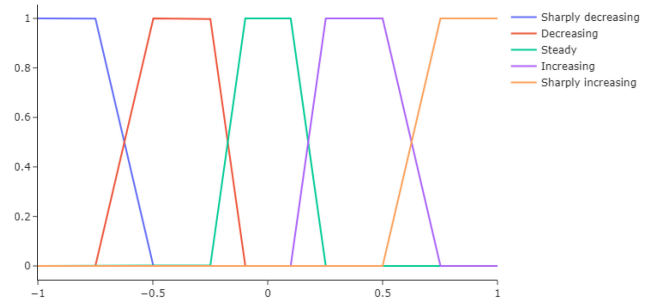


Fig. 6. Membership functions for segment trend summarizers.

better characteristics, a more in-depth study is needed in future work.

B. Data Labeling and Feature Discovery

While data characterization and simplification in TS have been primarily focused on a geometrical point of view as yet, experts face the significant challenge of identifying patterns or features existing in TS. Even so, medical staff play a fundamental role in defining their preferences for events' identification, which are later needed to be communicated in NL. FS theory is then employed to label the different variables, enabling a qualitative description of quantitative measures in the different features found within the domain. At this point, it seems convenient to remark that there is not a necessity of an existing labeled data set, since every single TS is characterized on demand. Consequently, this involves the computation of different summarizers (linguistic terms) which are a source of interest to the expert and are exposed subsequently.

- 1) *Glucose Level Summarizers*: Thresholds in the measurement of glucose are defined to determine the state of a patient, considering that each registered value x_t must be accompanied with a linguistic term, such as *very high*, *high*, *medium*, *low*, or *very low*. Actually, the different labels may differ depending on every case of study, but have been approximated according to the predefined intervals provided by ADA [18] (see Fig. 5).

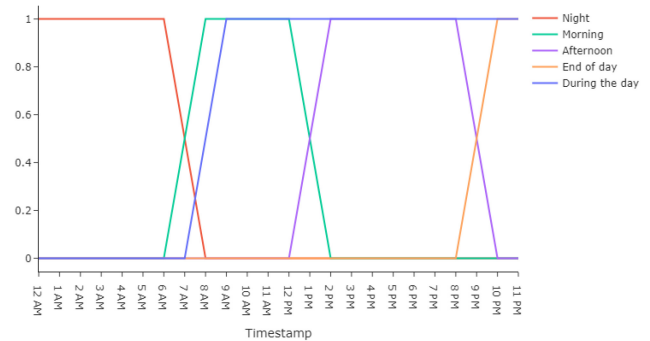


Fig. 7. Membership functions for day moment summarizers.

- 2) *Segment Trend Summarizers*: The slope m present in a segment between times t_1 and t_2 is calculated intending to estimate the variation between the two samples (e.g., either *rapidly*, *increasing* or *decreasing*, and *steady*) as specified in Fig. 6.
- 3) *Day Moment Summarizers*: Enriching the linguistic summary is a must in an attempt to clearly point out the time period where the set of samples have been registered. The ones considered by our system are being referred as *in the morning*, *during the afternoon*, *at night* or *at the end of the day* (see Fig. 7).

Once the summarizers have been successfully applied, type 1 protoforms can be established (e.g., *Glucose level is high* or

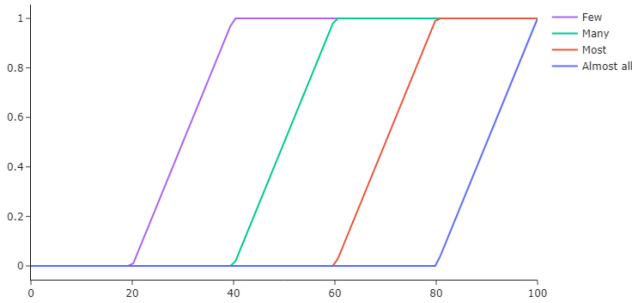


Fig. 8. Membership functions for quantifiers.

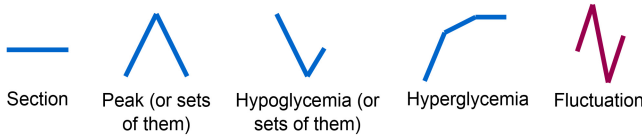


Fig. 9. Common patterns identified in the system.

Trend values are steady). In the task to comprehend all points available in a segment, the need of using a *fuzzy quantifier* Q emerges as a versatile tool to aggregate them all; every feature fitting into a certain summarizer is considered when generalizing the collection of samples that satisfy a given condition within all identified values in the domain (1) [30]. Those considered in the system (Fig. 8) are *few*, *many*, *most*, and *almost all*; essential terms for the construction of type 2 protoforms (e.g., *Almost all glucose levels are medium*)

$$\tau(QA \text{ are } S) = \mu_Q \left(\frac{1}{n} \sum_{i=1}^n \mu_S(a_i) \right). \quad (1)$$

With the purpose of generating protoforms in different intervals, a qualifier R is considered aiming to obtain a subset of the whole data. Therefore, the final evaluated protoforms have the form $R Q A \text{ are } S$ (e.g., *At night most glucose levels are high*). This way, all possible combinations are computed but a post-processing criterion is used to avoid those where $\text{DoT} < 0.7$, which is defined as

$$\text{DoT} = \mu_Q \left(10^2 \left(\frac{\sum_{i=1}^n \mu_R(y_i) \wedge \mu_S(y_i)}{\sum_{i=1}^n \mu_R(y_i)} \right) \right) \quad (2)$$

where $\wedge : [0, 1] \rightarrow [0, 1]$ denotes minimum and $\mu_Q, \mu_R, \mu_S : \mathbb{R} \rightarrow [0, 1]$ refer to the membership functions of the quantifier, the qualifier, and the summarizer, respectively, considering every single sample within the obtained subset.

Concurrently, simple geometrical structures characterization give rise to discover common patterns or key features that point out the importance of the underlying data, in a way that the description is provided with a higher level of semantic by the compactness of the progression of the variable in the system. These high-level features are exposed in Fig. 9 and are critical for the generation of NL in the process of establishing a detailed perspective of the TS.

As can be seen, *sections* (individual segments) do not only enrich data description but also refer to the different segments used for the TS characterization. Then, their main purpose is to be the basis to the construction of high-level structures such

as a *peak* p which is composed of two segments s_1 , whose trend is labeled as increasing or sharply increasing, and s_2 , whose trend label is (sharply) decreasing. The glucose level summarizer is computed with the point $(s_1 \cap s_2)$. Therefore, s_1 and s_2 are no longer considered sections in the system, as they start to belong to a higher abstraction level. Afterward, next analysis pretends to capture *sets of peaks* when appearing consecutively ($p_1 \cap p_2 \neq \emptyset$). Reached this point, the process concludes as no extra features are available to discover. After that, the remaining summarizers are reevaluated considering all points located within the partial sections that shape the specific pattern. The rest of the relevant phenomena in the domain of the problem are described next.

- 1) *Hypoglycemia*: Either a valley where the lowest point is labeled as *low* or *very low*, or one or more sections, where the first one has either its first or last point labeled as *low* or *very low*, being the subsequent samples labeled the same.
- 2) *Hyperglycemia*: Different from a peak as when a high or very high value is reached, the next values are labeled the same way in a prolonged period of time.
- 3) *Fluctuation*: It must be noted that this kind of event is present along the whole TS, but it is only considered by the time a peak, labeled as (very) high, is followed by a valley (or hypoglycemia episode) or vice-versa, being described this way in the final linguistic summary.

Notice that hypoglycemia and hyperglycemia episodes are characterized with a major importance as they refer to situations that may cause a harmful damage to the patient, in contrary to what may happen with isolated peaks, as they may be reflecting a normal scenario in the user's daily life (e.g., food ingestion or physical effort).

C. Linguistic Summary as Glucose Levels Descriptor

As a result of the previous section, it emerges the need to compute human capability to extrapolate the modeled knowledge into statements, via the utilization of a pre-established template (as shown in Table I). This template has been designed to summarize the set of protoforms generated in the analytical process, which for space reasons could not be included here. The preferences in terms of the generated text are oriented to the target users (TUs) (healthcare professionals and diabetics or their relatives) and based on expert guidelines, involving the linguistic context itself as well. Regrettably, including all the information obtained from the analyzed protoforms may result in a repetitive summarization of the TS, necessitating the development of a quality framework to selectively choose those that conform to predefined rules (see Table II). In summary, these rules state: Rule 1 establishes that abstract representations, such as peaks or fluctuations, are included before simpler descriptions. Rules 2 and 11 are related to make the summary shorter. The same purpose has rules 4, 9 and 13, that remove sentences that represent superfluous information to the final user. Rule 3 sets the threshold of 0.7 as the DoT to be accomplished by the generated linguistic expressions. This measure has been established heuristically under the supervision of experts, in order to filter only relevant

TABLE I
 LINGUISTIC SUMMARY TEMPLATE

| Template | Description |
|---|---|
| Collected glucose levels (GL) are between [hh]:[mm] /am /pm and [hh]:[mm] /am /pm /./ /., and a [quantity] amount of data has been lost /./ /.. Due to this lost rate, a reliable summary cannot be given.] | Quantity: minor, moderate, high |
| /User has experienced [quantifier] [glucose level] GLs [day moment].] /User GLs were normal during the day.] /User GLs present different fluctuations along the day.] /[day moment] /Next /Then /Moreover /Consecutively /Subsequently /Later on /At last [.] | Glucose level: very low, low, medium, high, very high Quantifier: almost all, most, many, few Day moment (as a connector): during the whole day, in the morning, during the afternoon, at night, at the end of the day |
| /A /An [event] of [value] mg/dL has been detected at around [hh]:[mm] /am /pm /./ /., followed by a [event] of [value] mg/dL at [hh]:[mm] /am /pm.] /and lasted for [period] /hours /minutes.]] /[quantifier] GLs have been [glucose level] during [period] /hours /minutes with a [trend] trend.] | Glucose level: very low, low, medium, high, very high Event: peak, hypoglycemia episode Quantifier: almost all, most, many, few Trend: sharply decreasing, decreasing, steady, increasing, sharply increasing Value: exact registered glucose level Period: duration |
| /A set of [number] [event] have appeared for around [period] /hours /minutes with a /maximum /minimum of [value] mg/dL at around [hh]:[mm] /am /pm.] /An hyperglycemia episode with a maximum of [value] mg/dL has been registered at around [hh]:[mm] /am /pm and lasted for [period] /hours /minutes.] | Event: peaks, hypoglycemia episodes Value: exact registered glucose level Period: duration |

information and reduce the size of the response. The rules that modify the output format are Rules 5, 8, and 12, that change some words of the final summary to make it clearer. Rules 6 and 7 simplify the representation of time and Rule 10 is included to generate a fluent discourse.

This phase terminates with the acquisition of an NL summarized text that complies with the modeled knowledge and the quality requirements established by the end users.

V. RESULTS AND CONSIDERATIONS

A. Experimental Results

The proposed linguistic summarization approach for continuous glucose monitoring data description has been proved effective when tested on a particular data set [53] that has been obtained from a diabetic patient. It is noteworthy to stress that collected data has been securely saved within a database that does not contemplate the inclusion of any personal information. Notwithstanding, the different procedures involved in this research have been performed in accordance to the standards set forth the 1964 Helsinki declaration and its latter updates and refinements.

 TABLE II
 QUALITY FRAMEWORK RULES FOR LINGUISTIC SUMMARY GENERATION

| # | Rule |
|----|--|
| 1 | The relevance of the protoforms is noted performing a top-down description. |
| 2 | The final number of sentences must be the lowest possible. |
| 3 | The degree of truth (DoT) of the generated expressions must be greater or equal than 0.7. |
| 4 | Protoforms of day intervals containing <i>few/many</i> quantifiers when describing <i>medium</i> glucose levels are not considered. Exceptional case: quantifier <i>many</i> describing <i>medium</i> glucose levels <i>during the whole day</i> . |
| 5 | <i>Medium</i> levels are referred to as normal glucose levels. |
| 6 | Shown hours are approximated to the closest quarter and the term <i>around</i> is included before this data. |
| 7 | Event duration is approximated to the closest number of hours. If less than one hour, minutes are shown. |
| 8 | Description must emphasize distinctive phenomena. |
| 9 | Expressions with time intervals involving less than 1 hour are not included. |
| 10 | Connectors are variable along the description. Last sentence begins with <i>At last</i> connector. |
| 11 | Partial consecutive summaries with similar information are aggregated. |
| 12 | The name of the variable glucose level in the domain of the problem is only typed this way at the beginning; latter allusions use the acronym GL. |
| 13 | If a whole day protoform indicates normal glucose levels, the rest of day moments protoforms are not included in the final summary. |
| 14 | If no protoform is activated for a time interval, the system indicates its variability because of different fluctuations of data. |

The specific case studies that are considered here, in terms to demonstrate the viability of the developed expert-guided system, concern the following days: 19 September 2023, 24 September 2023, and 3 November 2023 (see Table III).

First, it is relevant to highlight that, as specified in [25], the threshold of 0.2 used for the TS segmentation has demonstrated optimal performance in streamlining the TS, eliminating redundancies and preserving its fundamental shape, serving as the input data for the algorithm.

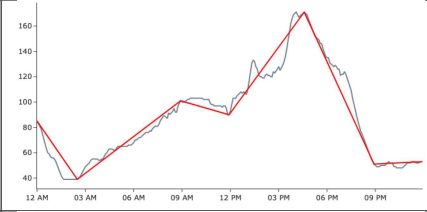
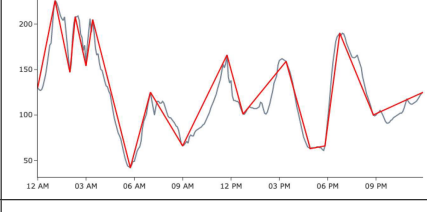
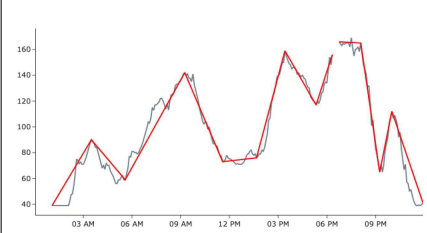
Additionally, the protoforms that the system has initiated (DoT > 0) based on the input data are shown in Table IV in the form $R Q A$ are S . The first digit of the protoform ID in this Table represents the number of the TS in Table III. Therefore, every single TS in this proposal generates 100 different protoforms to be computed as calculated in (3), where n is the total number of elements of each set, r refers to the amount of them (subset) taken into the computation of each protoform, and for each qualifier R , referring to the day moment, each quantifier Q undergoes evaluation for every individual glucose level labeled through a summarizer S

$$\zeta = \frac{n_R!}{r_R! \cdot (n_R - r_R)!} \cdot \frac{n_Q!}{r_Q! \cdot (n_Q - r_Q)!} \cdot \frac{n_S!}{r_S! \cdot (n_S - r_S)!} \quad (3)$$

Perceive that only those exceeding the minimum required DoT of 0.7 are accepted within the system and, their final inclusion in the linguistic summary corresponds to their surpassing of the quality rules specified in Table II. Despite this, all protoforms have been calculated as specified in (2), which may be visually contemplated in Fig. 10 for protoform (p.) 312, detailed in Table IV.

As depicted in Fig. 10, the graph illustrates the membership degrees for the summarizer S (*high*) and the qualifier

TABLE III
LINGUISTIC DESCRIPTION OF DIFFERENT TS

| # | TS | Summary | Exec. Time |
|---|---|---|------------|
| 1 |  | Collected glucose levels (GLs) are between 12:00 am and 11:55 pm. User GLs present different fluctuations along the day. At night, an hypoglycemia episode of 39.0 mg/dL has been detected at around 02:30 am. In the morning, most GLs have been normal during 8 hours with a sharply increasing trend. Then, a peak of 171.0mg/dL has been detected at around 04:30 pm. User has experienced most very low GLs at the end of the day. At last, an hypoglycemia episode with a minimum of 48.0 mg/dL has been detected at around 09:00 pm and lasted for 3 hours. | 1.7s |
| 2 |  | Collected glucose levels (GLs) are between 12:00 am and 11:55 pm. User GLs were normal during the day. At night, a set of 3 peaks have appeared for around 6 hours with a maximum of 226.0 mg/dL at 01:00 am. Next, an hypoglycemia episode of 42.0 mg/dL has been detected at around 05:45 am. During the afternoon, an hypoglycemia episode of 63.0 mg/dL has been detected at around 05:00 pm followed by a peak of 190.0 mg/dL at around 06:45 pm. At last, almost all GLs have been normal during 3 hours with an increasing trend. | 2.1s |
| 3 |  | Collected glucose levels (GLs) are between 12:05 am and 11:55 pm, and a minor amount of data has been lost. User has experienced few high GLs during the whole day. User has experienced few low GLs at night. Then, an hypoglycemia episode of 39.0 mg/dL has been detected at around 01:00 am. Next, an hypoglycemia episode of 59.0 mg/dL has been detected at around 05:30 am. In the morning, a peak of 142.0mg/dL has been detected at around 09:15 am followed by an hypoglycemia episode of 73.0 mg/dL at around 11:30 am. User has experienced many high GLs during the afternoon. Consecutively, a peak of 159.0 mg/dL has been detected at around 03:30 pm. Subsequently, an hyperglycemia episode with a maximum of 169.0 mg/dL has been registered at around 05:15 pm and lasted for 3 hours. Later on, an hypoglycemia episode of 65.0 mg/dL has been detected at around 09:15 pm. User has experienced few very low GLs at the end of the day. At last, an hypoglycemia episode with a minimum of 39.0 mg/dL has been detected at around 11:45 pm. | 2.59s |

R (afternoon). Effectively, the membership function for *R* maintains a static profile as defined in Fig. 7, while the membership function for *S* is variant, but accurately capturing pertinent phenomena outlined in the linguistic summary. Specifically, values do increase by the time the peaks were detected at 09:15 A.M. and 03:30 P.M., respectively, plus the hyperglycemia episode happening during the afternoon. It is noteworthy that only the latter two events were factored into the computation, as the first one is not happening during *R*. Finally, the DoT determined quantifier *many* as the best descriptor.

Aiming to clarify possible misunderstandings, it is important to note how p. 105 is not included in the summary to satisfy rule 11 (previously described in Table II). Moreover, p. 205, p. 206, and p. 208 are neither shown as a consequence of p. 203 and rule 13. Also, rule 4 causes p. 106, p. 301, p. 309, and p. 315 not to appear neither.

According to the definition of the different quantifiers in the system, the activation of protoforms with more restrictive quantifiers, implies the activation of others that are less extensive to the observed values, i.e., if most glucose values are high, then many are high too. Such protoforms are excluded from Table IV when an upper quantifier exceeds the DoT of 0.7. Likewise, the system ignores its computation given the existence of a better descriptor, but they exist if examined analytically. Hence, a total of 34 protoforms is shown, whereas the total number of activated ones for all TS is 50. Nonetheless, 27 out of 100 unique protoforms have been studied within the examples provided.

To validate the proposed methodology in terms of the variety of the generated protoforms, the computation of the whole data set [53] (composed of 226 days¹) has been carried out.² It has been observed that around 89% of possible protoforms are activated, indicating a high diversity in the performed data sampling. In addition, approximately 60% of protoforms were finally evaluated by the system. Consequently, this high activation rate and the selective evaluation mechanism ensure how the system focuses on the most relevant and informative protoforms, making it highly efficient in scenarios where computational resources and time may be limited. Furthermore, thorough the TS exposed in this work, it has been demonstrated the system's robustness and versatility in capturing various data patterns at a time. Altogether, these findings underscore the method's applicability and effectiveness in real-world scenarios.

The assessment of the generated linguistic summaries has been conducted by the employment of the Degree of Usefulness (DoU), where, after normalization, $DoU \in [0, 1]$. This metric quantitatively measures the utility of the generated sentences from the viewpoint of the TUs, comprising diabetic patients, their relatives, and healthcare professionals; the rationale behind this decision is to determine the interpretability of the system, as guided by a singular expert, thereby necessitating the identification of its strengths and weaknesses

¹This data set is available at <https://zenodo.org/records/10713570>.

²Additional illustrative examples are available at <https://asia.ujaen.es/material/expanded-protoform-evaluation/index.html>.

TABLE IV
PROTOFORMS GENERATED BASED ON INPUT DATA

| ID | Protoform | DoT | System Approval | Final Acceptance |
|-----|--|------|-----------------|------------------|
| 101 | During the whole day, many GLs were medium. | 0.47 | | |
| 102 | During the whole day, few GLs were high. | 0.01 | | |
| 103 | At night, many GLs were very low. | 0.27 | | |
| 104 | At night, many GLs were low. | 0.38 | | |
| 105 | In the morning, almost all GLs were medium. | 0.78 | ✓ | |
| 106 | During the afternoon, few GLs were medium. | 0.81 | ✓ | |
| 107 | During the afternoon, many GLs were high. | 0.04 | | |
| 108 | At the end of the day, almost all GLs were very low. | 0.84 | ✓ | ✓ |
| 201 | During the whole day, few GLs were very low. | 0.01 | | |
| 202 | During the whole day, few GLs were high. | 0.02 | | |
| 203 | During the whole day, many GLs were medium. | 0.92 | ✓ | ✓ |
| 204 | At night, few GLs were high. | 0.23 | | |
| 205 | At night, few GLs were very high. | 0.95 | ✓ | |
| 206 | In the morning, many GLs were medium. | 1 | ✓ | |
| 207 | In the morning, most GLs were medium. | 0.54 | | |
| 208 | During the afternoon, few GLs were medium. | 0.89 | ✓ | |
| 209 | During the afternoon, few GLs were high. | 0.44 | | |
| 210 | At the end of the day, almost all GLs were medium. | 0.8 | ✓ | ✓ |
| 301 | During the whole day, few GLs were medium. | 0.95 | ✓ | |
| 302 | During the whole day, few GLs were high. | 0.84 | ✓ | ✓ |
| 303 | During the whole day, few GLs were very high. | 0.03 | | |
| 304 | At night, few GLs were very low. | 0.39 | | |
| 305 | At night, few GLs were low. | 1 | ✓ | ✓ |
| 306 | At night, many GLs were low. | 0.07 | | |
| 307 | At night, few GLs were medium. | 0.54 | | |
| 308 | In the morning, few GLs were low. | 0.15 | | |
| 309 | In the morning, many GLs were medium. | 0.72 | ✓ | |
| 310 | In the morning, few GLs were high. | 0.27 | | |
| 311 | During the afternoon, few GLs medium. | 0.52 | | |
| 312 | During the afternoon, many GLs were high. | 0.83 | ✓ | ✓ |
| 313 | At the end of the day, few GLs were very low. | 0.71 | ✓ | ✓ |
| 314 | At the end of the day, few GLs were low. | 0.04 | | |
| 315 | At the end of the day, few GLs were medium. | 1 | ✓ | |
| 316 | At the end of the day, many GLs were medium. | 0.13 | | |

from the standpoint of diverse stakeholders, as well as aiming to discern the different expectations and requirements of them, seeking to pinpoint the absent information for potential

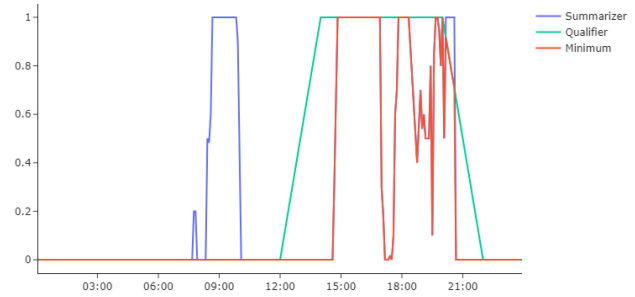


Fig. 10. Example of one protoform computation (p. 312).

TABLE V
EVALUATION OF THE RESULTS

| # | Statement | DoU | STDEV |
|----|---|------|-------|
| 1 | The linguistic summary provides all relevant information with sufficient precision and detail. | 0.73 | 0.3 |
| 2 | The explanations involve the majority of registered instances. | 0.79 | 0.22 |
| 3 | There are no inconsistencies found in the summary. | 0.81 | 0.32 |
| 4 | The explanation facilitates a causality understanding. | 0.64 | 0.19 |
| 5 | The linguistic summary does not include irrelevant information or incorrect statements. | 0.81 | 0.37 |
| 6 | The data is described progressively in time, with a useful inclusion of day time indicators for relevant phenomena description. | 0.82 | 0.51 |
| 7 | There is no need for support to assume the given statements. | 0.64 | 0.48 |
| 8 | The data is described without redundancy and briefly, understanding it very quickly. | 0.75 | 0.3 |
| 9 | There is no need for more expert-guided explanations. | 0.62 | 0.19 |
| 10 | It is not a must to change the expressions used in the sentences. | 0.63 | 0.26 |

inclusion in future approaches. The evaluation is conducted by ascribing a score on a 1–5 scale (from strongly disagree to strongly agree) for predetermined statements that can be seen in Table V accompanied by the mean-obtained results from a battery of linguistic summaries, as well as the standard deviation (STDEV) for a more detailed analysis.

B. Discussion and Limitations

The overall DoU obtained from survey respondents is around 0.73, demonstrating the viability and validity of the developed proposal in its preliminary stage. Fig. 11 shows the results for every group of TUs. Taking a deeper look, those statements referring to the included information in the summary, the comprehension of the instances and the quality of the description (s. 1, 2, 3, 5, 6, and 8) have achieved positive results (DoU ≥ 0.7), some presenting little variation (s. 1, 3, 5, and 8). Concurrently, results for statement 6 demonstrate the need of including fuzzy summarizers related to day intervals, plus the optimal utilization of them as qualifiers when creating high-level protoforms. As expected, there is a lack of causality understanding in general (s. 4), due to the unlabeled data that has been collected, not contemplating notes corresponding to food ingestion or physical activity, what translates into

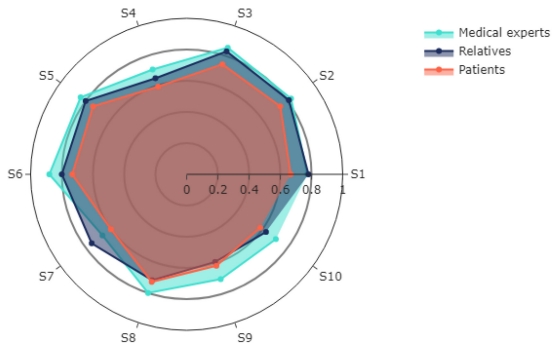


Fig. 11. Detailed DoU of the developed system.

necessitating more informed judgments with little changes in the expressions (s. 7, 9, and 10).

In detail, healthcare professionals show the highest acceptance of the system, emphasizing not only their positive view of the knowledge modeled (s. 5, 6, and 8) but also their support toward the no need of more expert-guided explanations (s. 9 and 10), even though this last statement corresponds to every individual's subjectivity. Patients and their relatives also have a positive perception of the system, although they are more reticent in general; expected in advance as the survey results represent an isolated case, despite the system's adaptability to individual patients.

It is noteworthy that our research has successfully presented a novel approach for diabetes care, diagnosis and treatment guided by expert knowledge in a multipatient ecosystem. However, since the collected data only rely on glucose levels measures, our approach has a limited scope as it could be important not to overlook essential factors that could impact over glucose management, i.e., meal timing, physical activity, medication usage, or insulin supply. Moreover, generalization cannot be applied as different patients may need to vary the system's parameters for glucose dynamics description; the proposed framework still considers the personalization of these parameters aiming to a betterment of the end-users' experience.

C. Quantitative Comparison Involving GPT-4

Given that the summarization methodology is unprecedented in the glucose domain, the performance of the artificial intelligence (AI) chatbot running GPT-4o technology has been tested regarding the steps-to-follow exposed in [25] together with the expanded considerations found in [24]. It is pretended to consider an overview of the results thrown by a nondedicated tool for linguistic summarization, carrying out a methodology including a discussion forum with technical experts when evaluating the generated outputs. This validation method is also done in [24], where the description of glucose TS is conducted in accordance with two different tasks. First, the proposal focus on evaluating metrics' generation and, in the second part, it creates a summarization of data highlighting the most relevant aspects; the knowledge is modeled according to expert guidance as well as the requirements the summary should meet in terms of content and structure.

Aiming to evaluate the performance of this proposal to describe one-day-glucose TS with GPT-4o, a battery of tests has been organized using three different scenarios for two different kinds of TS, the original and the segmented one (as shown in Table VI):

- 1) GPT-4o chatbot [54];
- 2) Data Analyst plug-in powered by GPT-4 [24];
- 3) our proposed GLiDTS-based system.

Note that data imputation is made over the collected TS as specified in our methodology.

The prompt starts providing GPT-4o with the file containing the TS to describe, followed by the next instructions: "Act as an endocrinologist expert in diabetes care and analyze the provided continuous glucose monitored data for its description including the most relevant information, such as anomalous events (hyperglycemia and hypoglycemia episodes), variation and summarized values. To do so, make a summary in NL contemplating day intervals." The generated output from the LLM for the three different TS illustrated in Table III are shown in Table VI. Note that superfluous information provided by this technology regarding potential adjustments in diabetes management, as shown in the first output, is avoided for space reasons.

To assess the performance of this technology, different criteria are defined to measure how closely GPT-4 systems meet anticipated human expectations. Table VII reflects these criteria together with the satisfaction degree of the previous generated summaries (Table VII) and the scoring for each one: 1 (satisfied) or 0 (nonsatisfied). For example, a score of 0-0-1 means that only the third of the three analyzed TS meets this criterion. It is essential to emphasize that quality directives are designed to address the needs of all expected audience, including patients, their relatives, and medical staff.

The obtained results have proven the viability and validity of the proposed GLiDTS-based system, after achieving a degree of acceptance of 86%. However, it must be highlighted the notable score of 76% the Data Analyst plug-in of GPT-4 has reached when analyzing the segmented TS; the process of simplifying the collected data has been verified to be feasible, as through the utilization of the two technologies, the analysis of the original TS makes the results obtained worse. In general, GPT-4 technology causes the summaries to have invalid statements due to its incapability for data aggregation (consecutive hypoglycemia events are expressed separately), and its inflexibility when modeling the expected knowledge (sometimes considering hypo/hyperglycemia episodes while sometimes not), generating confusion when looking at the graphical representation of the corresponding TS. The proposed system is capable of capturing all significant trends while GPT-4 does not show interest in them, while providing different statistical metrics the approach we present does not. In contrary, we have put the focus on a semantic output that GPT-4 sometimes contemplates. Nonetheless, in most cases, all technologies meet the requirements expected for insignificant measures omission, nonrequired data petition, relevant events' timing and the achievement of concise descriptions. Additionally, it is noticed that while the generated system in this work may expose little redundancies, GPT-4 approaches always

TABLE VI
GPT-4O AND DATA ANALYST PLUG-IN BY GPT-4 PERFORMANCE IN GLUCOSE TS SUMMARIZATION

| TS# | TS Type | Points# | Technology | Summary | Exec. Time |
|-----|-----------|---------|--------------------|---|------------|
| 1 | Original | 288 | GPT-4o | The continuous glucose monitoring data from September 19, 2023, shows an average glucose level of 89.4 mg/dL with a standard deviation of 36.1 mg/dL. Throughout the day, glucose levels ranged from a minimum of 39 mg/dL, which is a hypoglycemic episode, to a maximum of 171 mg/dL. No hyperglycemic events occurred. A total of 287 measurements were recorded, with the median glucose level at 88 mg/dL. Episodes of hypoglycemia were noted but specific details on frequency or duration would require further analysis. This summary provides an overview of the glucose level fluctuations and notable events for the day, useful for guiding further detailed examination and potential adjustments in diabetes management. | 9s |
| | | | GPT-4-Data Analyst | The continuous glucose monitoring data shows numerous instances of hypoglycemia, totaling 111 events throughout the day. The hypoglycemic episodes occur at various times, with glucose levels ranging from 39 mg/dL to 69 mg/dL, and an average of around 53 mg/dL. There are no instances of severe hyperglycemia (above 180 mg/dL) in the dataset. Notably, the data reveals regular fluctuations below the hypoglycemic threshold, particularly evident during late night and early morning hours. (Superfluous information) | 21s |
| | Segmented | 7 | GPT-4o | The data shows glucose measurements from September 19, 2023. On this day, the glucose levels ranged from a minimum of 39 mg/dL to a maximum of 171 mg/dL, with an average glucose level of approximately 84 mg/dL. There were 3 episodes of hypoglycemia, where glucose levels fell below 70 mg/dL. There were no episodes of hyperglycemia, as the glucose levels did not exceed 180 mg/dL. (Superfluous information) | 7.9s |
| | | | GPT-4-Data Analyst | The continuous glucose monitoring data from September 19, 2023, includes seven readings, with glucose levels averaging 84.3 mg/dL. Hypoglycemic events occurred three times, most notably early in the morning at 2:30 am and late at night at 23:55 pm. The lowest recorded glucose level was 39 mg/dL. There were no instances of hyperglycemia. The overall glucose variation was substantial, with a standard deviation of 44.7 mg/dL. | 18s |
| 2 | Original | 288 | GPT-4o | The continuous glucose monitoring data from September 24, 2023, reveals an average glucose level of 121.1 mg/dL with a variation of 41.8 mg/dL. There were 32 episodes of hyperglycemia, where glucose levels exceeded 180 mg/dL, and 34 episodes of hypoglycemia, with levels falling below 70 mg/dL. Notably, glucose levels fluctuated between a minimum of 42 mg/dL and a maximum of 226 mg/dL throughout the day. (Superfluous information) | 8.9s |
| | | | GPT-4-Data Analyst | The continuous glucose monitoring data reveals an average glucose level of 121.15 mg/dL with a variability of 41.75 mg/dL. The lowest recorded glucose was 42 mg/dL and the highest was 226 mg/dL. There were 34 hypoglycemic episodes, most commonly occurring around 5 pm, and 32 hyperglycemic episodes, most frequently at 1 am. (Superfluous information) | 19.2s |
| | Segmented | 17 | GPT-4o | The continuous glucose monitoring data from September 24, 2023, shows multiple episodes of glucose level fluctuations. Hyperglycemia was observed four times: at 01:05, 02:20, 03:25, and 18:45. Hypoglycemia occurred four times as well: at 05:45, 09:00, 16:55, and 17:50. The average glucose level was approximately 134 mg/dL, indicating significant variance throughout the day. (Superfluous information) | 9.1s |
| | | | GPT-4-Data Analyst | The glucose measurements varied significantly, with a minimum of 42 mg/dL, and a maximum of 226 mg/dL. Hypoglycemia occurred at 05:45, 09:00, 16:55, and 17:50, whereas hyperglycemia episodes were noted at 01:05, 02:20, 03:25, and 18:45. Overall, the mean glucose level was 133.94 mg/dL. (Superfluous information) | 23s |
| 3 | Original | 272 | GPT-4o | The continuous glucose monitoring data from November 3, 2023, indicates an average glucose level of 98.9 mg/dL with a standard deviation of 36.0 mg/dL, highlighting considerable day-to-day variability. The data comprises 271 measurements, spanning from a minimum of 39 mg/dL to a maximum of 169 mg/dL. Notably, there were no hyperglycemic events observed. The median glucose value was approximately 91.8 mg/dL. Episodes of hypoglycemia were recorded, which are critical for evaluating and possibly refining diabetes management plans. (Superfluous information) | 17.2s |
| | | | GPT-4-Data Analyst | The continuous glucose monitoring data contains 271 entries from November 3, 2023, sampled every 5 minutes. The glucose levels range from a low of 39 mg/dL, indicating episodes of severe hypoglycemia, to a high of 169 mg/dL, which could suggest periods of hyperglycemia. The average glucose concentration is approximately 99 mg/dL, with a standard deviation of about 36 mg/dL. Notable is the presence of hypoglycemic readings at the beginning of the dataset, with multiple values at 39 mg/dL. More detailed analysis regarding the specific timing of hyperglycemic peaks and other fluctuations throughout the day could provide further insights into glucose control and necessary adjustments in diabetes management for the individual. | 38.5s |
| | Segmented | 15 | GPT-4o | Analyzing the continuous glucose monitoring data from November 3, 2023, reveals that the individual's glucose levels fluctuated significantly throughout the day. There were 15 recorded measurements with an average glucose level of approximately 100 mg/dL. The standard deviation was high at 48 mg/dL, indicating variable glucose control. The data shows five episodes of hypoglycemia, with the lowest reading at 39 mg/dL. These hypoglycemic events were observed mostly during the early hours. No episodes of hyperglycemia were recorded; the highest glucose level was 166 mg/dL. This summary suggests a pattern of nocturnal hypoglycemia that needs attention to adjust diabetes management strategies effectively. | 12.3s |
| | | | GPT-4-Data Analyst | The data shows glucose measurements taken on November 3rd, 2023, ranging from 39 mg/dL to 166 mg/dL with an average of around 100 mg/dL. There were hypoglycemic events early in the morning between 00:05 am and 01:05 am. Hyperglycemic events occurred around 09:15 am and later in the day. The data exhibits significant variation, with glucose levels often swinging widely over a short period, suggesting episodes of acute glycemic changes. (Superfluous information) | 41.3s |

suggest vague recommendations accompanied by superfluous information. Also, GPT-4 cannot deal with a gradual summarization of data in a progressive manner over the course of

the day. Finally, in terms of execution time, while GPT-4o chatbot takes around 11 s and the GPT-4 Data Analyst Plug-In performance is prolonged up to 27 s on average, the

TABLE VII
PERFORMANCE OF GLiDTS AND GPT-4 IN SATISFYING EVALUATION CRITERIA

| # | Criterion | GLiDTS | GPT-4o-Original TS | GPT-4o-Segmented TS | GPT-4 Data Analyst-Original TS | GPT-4 Data Analyst-Segmented TS |
|----------------------|---|--------------|--------------------|---------------------|--------------------------------|---------------------------------|
| C1 | The text provided is not composed of invalid statements. | 1 | 0 | 0.33 (0-0-1) | 0.33 (0-0-1) | 0.33 (0-1-0) |
| C2 | The most relevant aspects of the TS have been highlighted. | 1 | 0.33 (0-1-0) | 1 | 1 | 1 |
| C3 | Hyperglycemia and hypoglycemia events are captured. | 1 | 1 | 1 | 1 | 1 |
| C4 | Sharp increases and decreases are detected in the collected values. | 1 | 0 | 0 | 0 | 0 |
| C5 | The values of most collected instances have been summarized in a semantic and a coherent manner. | 1 | 0.33 (0-1-0) | 0.67 (0-1-1) | 0.33 (0-0-1) | 0.67 (1-0-1) |
| C6 | The values of most collected instances have been summarized with statistical metrics. | 0 | 1 | 1 | 1 | 1 |
| C7 | Insignificant measures are omitted. | 0.67 (1-0-1) | 0.67 (1-1-0) | 0.67 (1-1-0) | 0.67 (1-1-0) | 1 |
| C8 | Relevant episodes are mentioned together with the day moment or the exact time of its occurrence. | 1 | 0 | 1 | 0.67 (1-1-0) | 1 |
| C9 | Irrelevant required data situations are not specified. | 1 | 1 | 1 | 0.67 (1-1-0) | 1 |
| C10 | The TS is not described redundantly. | 0.33 (0-1-0) | 0.67 (1-0-1) | 0.67 (1-0-1) | 0.33 (0-1-0) | 1 |
| C11 | The TS is described step-by-step over time. | 1 | 0 | 0 | 0 | 0.33 (0-0-1) |
| C12 | The generated natural language is not specialized in the domain. | 1 | 0.67 (0-1-1) | 1 | 0.67 (1-1-0) | 1 |
| C13 | Superfluous information is omitted. | 1 | 0 | 0.33 (0-0-1) | 0 | 0.33 (1-0-0) |
| C14 | A short and concise description is provided. | 1 | 0.33 (0-1-0) | 1 | 0.67 (1-1-0) | 1 |
| Overall score | | 0.86 | 0.45 | 0.69 | 0.52 | 0.76 |

proposed system lasts around 2 s to generate the linguistic summary.

The designed decision criteria in our work is centered in evaluating the linguistic description itself for its communication to the potential target audience these technologies could be integrated in, in contrast to [24], which specifies criteria for evaluation by experts; the expected output is predefined in the prompt using a template composed of acronyms and codifications, not focusing on generating a language that is familiar and close to the end users, deriving in the necessity of a higher mental effort. Additionally, along with the effort they put on stating the expert knowledge as an input, this proposal does not take advantage of the proper nature of LLMs. In the first case, the ability of GPT-4 to generate NL is avoided and, in the second case, the huge amount of data the technology has been trained with is eluded by the time the knowledge to model is defined.

In general, GPT-4 technology has demonstrated some useful and reasonable statements, despite both errors or hallucinations in identifying relevant phenomena, and mistakes in providing clinical conclusions or advice, due to an inappropriate code generation and/or its misinterpretation. For instance, little variability is assumed as a regular and proper control of insulin while all glucose levels are high as shown in [24]. At this point, it is highlighted how the provided prompt consists in a single petition, even though the output obtained from an LLM could be polished through an iterative process. However, identifying such mistakes or confusions requires an acute overview of the data-related cause-effect relationship with the generated output, not resolving the problem we are handling. Equally, the obtained output is stochastic, so the same TS analysis never generates an identical response, even with a template, and since the evaluation is based on the current model, the linguistic summary is not controllable over time.

Upon delving into the examination of the GPT-4 model generation process, it is important to emphasize the significant

computational resources needed, as specified in [55], requiring a supercomputer for complexity and scale reasons, prompting a discussion on the optimization of resource allocation to ensure environmental and economic sustainability. Therefore, it is accentuated the need for evaluating the GLiDTS-based system together with GPT-4 technology from a sustainable point of view, throughout their alignment with the United Nations Sustainable Development Goals (SDGs) [56]. Given their capability for being integrated in real word scenarios, both systems contribute to *SDG #3: Good Health and Well-Being*, as characterized to develop a new service in the healthcare domain for diabetes management. Nonetheless, there is a misalignment between GPT-4 technology and the goals *SDG #10: Reduced Inequalities* and *SDG #12: Responsible Consumption and Production*. This is a consequence of GPT-4 substantial monetary cost and significant resource consumption for its generation, which may exacerbate inequalities and lead to unsustainable resource utilization. Conversely, our GLiDTS-based system demonstrates a better alignment by effectively contributing to these goals, as it adheres to open science standards and is provided free of charge, as well as necessitating low resource consumption.

As a result, LLMs have shown that they can generate acceptable results, but the low interpretability and the great effort required for their construction, the low quality of the result, and its incapacity to be aligned with the SGDs have led us to discard its use. In this sense, this proposal has proven to be feasible, innovative, sustainable, and complete in both, a qualitative and a quantitative manner, demonstrating the need of developing specialized tools in the domain of the problem we are handling.

VI. CONCLUSION AND FUTURE WORK

This study has successfully introduced a complete proposal for the generation of high-quality linguistic summaries from

TS, establishing a knowledge representation model for a posterior generation of NL descriptions in the field of diabetes and the control of glucose measurements. To do that, a real-time multipatient monitoring architecture has been developed to keep track of the evolution of glucose levels, by means data acquired from an interstitial continuous glucose monitoring sensor.

The resulting multipatient framework allows to monitor and visualize diabetes measures in an interpretable way, thanks to the layer-based methodology, where experts and diabetic users (or their relatives) have been involved in its design. Consequently, this system tends to focus on relevant phenomena when observing a graphical representation of an individual's glucose data, considering different abstraction levels to effectively achieve this goal, in a similar way as humans do. Moreover, the use of the FL theory has been applied to extrapolate those key elements into words.

As future work, we aim to introduce a betterment of the linguistic descriptions by considering a more extended period of time, pretending to enable the assessment of a patient's historical data, while also considering the analysis of improvements or decays in eating habits.

REFERENCES

- [1] A. M. Egan and S. F. Dinneen, "What is diabetes?" *Medicine*, vol. 47, no. 1, pp. 1–4, 2019.
- [2] (Int. Diabetes Fed., Brussels, Belgium). *What is Diabetes*. Accessed: Dec. 12, 2023. [Online]. Available: <https://idf.org/about-diabetes/what-is-diabetes/>
- [3] V. Kalouguina and J. Wagner, "Challenges and solutions for integrating and financing personalized medicine in healthcare systems: A systematic literature review," *J. Risk Financ. Manag.*, vol. 13, no. 11, p. 283, Nov. 2020.
- [4] K. Rose, S. Eldridge, and L. Chapin, "The Internet of Things: An overview," *Internet Soc.*, vol. 80, no. 15, pp. 1–50, 2015.
- [5] D. Amghar and A. M. Chikh, "Extracting a linguistic summary from a medical database," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 12, p. 16, 2018.
- [6] J. Buelvas, D. Múnera, D. P. Tobón V, J. Aguirre, and N. Gaviria, "Data quality in IoT-based air quality monitoring systems: A systematic mapping study," *Water, Air, Soil Pollut.*, vol. 234, no. 4, p. 248, 2023.
- [7] N. Hossein Motlagh et al., "Unmanned aerial vehicles for air pollution monitoring: A survey," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21687–21704, Dec. 2023.
- [8] D. P. Rani, D. Suresh, P. R. Kapula, C. M. Akram, N. Hemalatha, and P. K. Soni, "IoT based smart solar energy monitoring systems," *Mater. Today, Proc.*, vol. 80, pp. 3540–3545, Jan. 2023.
- [9] Q. Zheng et al., "Smart contract-based agricultural service platform for drone plant protection operation optimization," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21363–21376, Dec. 2023.
- [10] L. P. Maguluri, J. Ananth, S. Hariram, C. Geetha, A. Bhaskar, and S. Boopathi, "Smart vehicle-emissions monitoring system using Internet of Things (IoT)," in *Handbook of Research on Safe Disposal Methods of Municipal Solid Wastes for a Sustainable Environment*. Hershey, PA, USA: IGI Global, 2023, pp. 191–211.
- [11] S. Chen, S. Zhao, and C. Huang, "An automatic malaria disease diagnosis framework integrating blockchain-enabled cloud-edge computing and deep learning," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21544–21553, Dec. 2023.
- [12] V. Salunkhe and R. Sujatha, "Secure sharing of healthcare data using a decentralized blockchain network for the Internet of Medical Things (IoMT)," *Int. J. Comput. Digit. Syst.*, vol. 14, no. 1, pp. 1–30, 2023.
- [13] S. Faccioli, B. Ozaşlan, J. F. Garcia-Tirado, M. Breton, and S. Del Favero, "Black-box model identification of physical activity in type-1 diabetes patients," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2018, pp. 3910–3913.
- [14] K. A. McMillan, A. Kirk, A. Hewitt, S. MacRury, and M. Lennon, "Methods for combining continuously measured glucose and activity data in people with type 2 diabetes: Challenges and solutions," *J. Rehabil. Assist. Technol. Eng.*, vol. 5, Jan.–Dec. 2018, Art. no. 2055668318782805.
- [15] C. Rodríguez-León, C. Villalonga, M. Muñoz-Torres, J. R. Ruiz, and O. Banos, "Mobile and wearable technology for the monitoring of diabetes-related parameters: Systematic review," *JMIR mHealth uHealth*, vol. 9, no. 6, 2021, Art. no. e25138.
- [16] T. Zhu, L. Kuang, J. Daniels, P. Herrero, K. Li, and P. Georgiou, "IoMT-enabled real-time blood glucose prediction with deep learning and edge computing," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 3706–3719, Mar. 2023.
- [17] O. S. Blebalu, "The fuzzy future: Embracing the potential of fuzzy functions," *Afr. J. Adv. Pure Appl. Sci.*, vol. 2, no. 2, pp. 103–115, 2023.
- [18] N. A. ElSayed et al., "6. Glycemic targets: Standards of care in diabetes—2023," *Diabetes Care*, vol. 46, no. 1, pp. 97–110, 2022.
- [19] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [20] S. Bhuvaneswari and K. Soujanya, "State of art in fuzzy logic," *J. Surv. Fish. Sci.*, vol. 10, no. 2S, pp. 3562–3582, 2023.
- [21] K. Liu et al., "Machine learning models for blood glucose level prediction in patients with diabetes mellitus: Systematic review and network Meta-analysis," *Med. Inform.*, vol. 11, no. 1, 2023, Art. no. e47833.
- [22] R. Rastogi and M. Bansal, "Diabetes prediction model using data mining techniques," *Meas., Sens.*, vol. 25, Feb. 2023, Art. no. 100605.
- [23] S. Langarica, M. Rodríguez-Fernandez, F. J. Doyle, and F. Núñez, "A probabilistic approach to blood glucose prediction in type 1 diabetes under meal uncertainties," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 10, pp. 5054–5065, Oct. 2023.
- [24] E. Healey, A. Tan, K. Flint, J. Ruiz, and I. Kohane, "Leveraging large language models to Analyze continuous glucose monitoring data: A case study," *medRxiv*, submitted for publication.
- [25] C. Martínez-Cruz, J. F. Gaitán Guerrero, J. L. López Ruiz, A. J. Rueda, and M. Espinilla, "A first approach to the generation of linguistic summaries from glucose sensors using GPT-4," in *Proc. 15th Int. Conf. Ubiquitous Comput. Ambient Intell. (UCAmI)*, 2023, pp. 33–43.
- [26] B. Najafi, D. Horn, S. Marclay, R. T. Crews, S. Wu, and J. S. Wrobel, "Assessing postural control and wearal control strategy in diabetes patients using innovative and wearable technology," *J. Diabetes Sci. Technol.*, vol. 4, no. 4, pp. 780–791, 2010.
- [27] V. R. Reddy, A. D. Choudhury, P. Deshpande, S. Jayaraman, N. K. Thokala, and V. Kaliaperumal, "DMSense: A non-invasive diabetes mellitus classification system using photoplethysmogram signal," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, 2017, pp. 71–73.
- [28] L. Zadeh, "Fuzzy logic = computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103–111, May 1996.
- [29] R. R. Yager, *On Linguistic Summaries of Data, Knowledge Discovery in Databases*, G. Piattetsky-Shapiro and B. Frawley, Eds., Cambridge, MA, USA: MIT Press, 1991, pp. 347–363.
- [30] R. Yager, "A new approach to the summarization of data," *Inf. Sci.*, vol. 28, no. 1, pp. 69–86, 1982.
- [31] L. A. Zadeh, "From computing with numbers to computing with words. from manipulation of measurements to manipulation of perceptions," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 46, no. 1, pp. 105–119, Jan. 1999.
- [32] P. Mehra, T. Seth, and P. K. Muhuri, "Generating quality IF-THEN rules for diabetes using linguistic summarization," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, 2020, pp. 1–7.
- [33] A. R. Nasser et al., "IoT and cloud computing in health-care: A new wearable device and cloud-based deep learning algorithm for monitoring of diabetes," *Electronics*, vol. 10, no. 21, p. 2719, 2021.
- [34] N. Marín and D. Sánchez, "On generating linguistic descriptions of time series," *Fuzzy Sets Syst.*, vol. 285, pp. 6–30, Feb. 2016.
- [35] Ü. Kaya, D. Akay, and S. Ş. Ayan, "EEG-based emotion recognition in neuromarketing using fuzzy linguistic summarization," *IEEE Trans. Fuzzy Syst.*, early access, Apr. 22, 2024, doi: [10.1109/TFUZZ.2024.3392495](https://doi.org/10.1109/TFUZZ.2024.3392495).
- [36] A. Cascallar-Fuentes, J. Gallego-Fernández, A. Ramos-Soto, A. Saunders-Estévez, and A. Bugarín-Diz, "Automatic generation of textual descriptions in data-to-text systems using a fuzzy temporal ontology: Application in air quality index data series," *Appl. Soft Comput.*, vol. 119, Apr. 2022, Art. no. 108612.

- [37] A. Cascallar-Fuentes, A. Ramos-Soto, and A. Bugarín, “Meta-heuristics for generation of linguistic descriptions of weather data: Experimental comparison of two approaches,” *Fuzzy Sets Syst.*, vol. 443, pp. 173–202, Aug. 2022.
- [38] S. Genc, D. Akay, F. E. Boran, and R. R. Yager, “Linguistic summarization of fuzzy social and economic networks: An application on the international trade network,” *Soft Comput.*, vol. 24, pp. 1511–1527, Jan. 2020.
- [39] M. Peláez-Aguilera, M. Espinilla, M. Fernández, and J. Medina, “Fuzzy linguistic protoforms to summarize heart rate streams of patients with ischemic heart disease,” *Complexity*, vol. 2019, no. 1, 2019, Art. no. 2694126.
- [40] A.-P. Albín-Rodríguez, A.-J. Ricoy-Cano, Y.-M. de-la Fuente-Robles, and M. Espinilla-Estévez, “Fuzzy protoform for hyperactive behaviour detection based on commercial devices,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 18, p. 6752, 2020.
- [41] A. Jain, M. Popescu, J. Keller, M. Rantz, and B. Markway, “Linguistic summarization of in-home sensor data,” *J. Biomed. Inform.*, vol. 96, Aug. 2019, Art. no. 103240.
- [42] C. Martínez-Cruz, A. J. Rueda, M. Popescu, and J. M. Keller, “New linguistic description approach for time series and its application to bed restlessness monitoring for eldercare,” *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 4, pp. 1048–1059, Apr. 2022.
- [43] M. L. Elias, J. Burshtein, and V. R. Sharon, “OpenAI’s GPT-4 performs to a high degree on board-style dermatology questions,” *Int. J. Dermatol.*, vol. 63, no. 1, pp. 73–78, 2024.
- [44] Z. Kanjee, B. Crowe, and A. Rodman, “Accuracy of a generative artificial intelligence model in a complex diagnostic challenge,” *Jama*, vol. 330, no. 1, pp. 78–80, 2023.
- [45] “Freestyle libre 3 sensor.” Abbot. Accessed: Dec. 12, 2023. [Online]. Available: <https://www.freestyle.abbott/es-es/productos/freestylelibre-3.html>
- [46] “Nightscout xDrip+.” Nightscout contributors. Accessed: Dec. 12, 2023. [Online]. Available: <https://github.com/NightscoutFoundation/xDrip>
- [47] “Nightscout RESTful service framework.” Nightscout contributors. Accessed: Dec. 12, 2023. [Online]. Available: <https://github.com/cosmonaut/nightscout/blob/develop/swagger.yaml>
- [48] N. Farsi, N. Mahjouri, and H. Ghasemi, “Breakpoint detection in non-stationary runoff time series under uncertainty,” *J. Hydrol.*, vol. 590, Nov. 2020, Art. no. 125458.
- [49] A. Ali and J. K. Aggarwal, “Segmentation and recognition of continuous human activity,” in *Proc. IEEE Workshop Detect. Recognit. Events Video*, 2001, pp. 28–35.
- [50] Y. Tang, L. Zhang, Q. Teng, F. Min, and A. Song, “Triple cross-domain attention on human activity recognition using wearable sensors,” *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 5, pp. 1167–1176, Oct. 2022.
- [51] A. Saalfeld, “Topologically consistent line simplification with the Douglas-Peucker algorithm,” *Cartogr. Geogr. Inf. Sci.*, vol. 26, no. 1, pp. 7–18, 1999.
- [52] R. Castillo-Ortega, N. Marín, C. Martínez-Cruz, and D. Sánchez, “Linguistic comparison of time series using the end-point fit algorithm,” in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2015, pp. 1–8.
- [53] J. F. Gaitan-Guerrero, J. L. Lopez Ruiz, C. Martínez-Cruz, and M. Espinilla, 2023, “T1GDUJA: Glucose dataset of a patient with type 1 diabetes mellitus,” Dataset, Kaggle. [Online]. Available: <https://zenodo.org/records/10713570>
- [54] “GPT-4o.” OpenAI. Accessed: May 5, 2024. [Online]. Available: <https://openai.com/index/hello-gpt-4o/>
- [55] “GPT-4.” OpenAI. Accessed: May 5, 2024. [Online]. Available: <https://openai.com/index/gpt-4/>
- [56] (United Nat., New York, NY, USA). *Take Action for the Sustainable Development Goals—United Nations Sustainable Development*. Accessed: May 5, 2024. [Online]. Available: <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>



Juan Francisco Gaitán-Guerrero was born in Bailén, Jaén, Spain, in 2000. He received the B.Sc. degree in computer science from the University of Jaén, Jaén, where he is currently pursuing the M.Sc. degree in computer science.

He is currently a member of the Advances in Intelligent Systems and Applications (ASIA) Research Group, University of Jaén, as a Research Assistant, primarily focusing on the Internet of Things and artificial intelligence applied in health environments.



José Luis López Ruiz was born in Linares, Jaén, Spain, in 1993. He received the M.Sc. and Ph.D. degrees in computer science from the University of Jaén, Jaén, in 2020 and 2023, respectively.

He is currently an Assistant Professor with the Department of Computer Science, University of Jaén. His current research interests are Internet of Medical Things, ambient assisted living, indoor location systems, and sustainability.



Macarena Espinilla was born in Jaén, Spain, in 1983. She received the M.Sc. and Ph.D. degrees in computer science from the University of Jaén, Jaén, in 2006 and 2009, respectively.

She currently serves as a Full Professor with the Computer Science Department, University of Jaén, where she is the responsible leader of the Research Group Avances en Sistemas Inteligentes y Aplicaciones (ASIA), which focuses on advancements in intelligent systems and applications. She has made significant contributions to the field of activity recognition for improving the quality of life for individuals. She has published over 50 papers in SCI-indexed journals. Her research interests primarily revolve around eHealth, ambient intelligence, ambient assisted living, and activity recognition.

activity recognition for

Prof. Espinilla received the Extraordinary Award for her Ph.D. in 2015. In 2022, her remarkable achievements led her to receive a medal in the Young Researchers Awards by the Royal Academy of Engineering in Spain. She serves as an Editorial Board Member for the *Journal of Universal Computer Science*, *IEEE ACCESS*, and the *Journal of Sensor and Actuator Networks*.



Carmen Martínez-Cruz received the Ph.D. degree in computer science and artificial intelligence from the University of Granada, Granada, Spain, in 2008.

She joined the University of Jaén, Jaén, Spain, in 2005, where she is currently an Associate Professor with the Computer Science Department. Her research has focused on the area of knowledge representation, fuzzy databases, ontologies, fuzzy logic, linguistic summarization, and eldercare technologies in which she has published a significant number of papers.