

# PROPOSAL FOR A METHODOLOGY TO CREATE AN ARTIFICIAL PANCREAS FOR THE CONTROL OF TYPE 1 DIABETES USING MACHINE LEARNING AND AUTOMATED INSULIN DOSING SYSTEMS (AID)

Rodrigo Vilanova<sup>1</sup> and Anderson Jefferson Cerqueira<sup>2</sup>

<sup>1</sup>Avantsec, Brasilia, Brazil

<sup>2</sup>Computer Science Department, University of Brasília (UnB), Brasilia, Brazil

## **ABSTRACT**

*The number of children, adolescents, and adults living with diabetes increases annually due to the lack of physical activity, poor diet habits, stress, and genetic factors, and there are greater numbers in low-income countries. Therefore, the aim of this article is to present a proposal for a methodology for developing a pancreas using artificial intelligence to control the required doses of insulin for a patient with type 1 diabetes (T1D), according to data received from monitoring sensors. The information collected can be used by physicians to make medication changes and improve patients' glucose control using insulin pumps for optimum performance. Therefore, using the model proposed in this work, the patient is offered a gain in glucose control and, therefore, an improvement in quality of life, as well as in the costs related to hospitalization.*

## **KEYWORDS**

*Machine Learning, Artificial Intelligence, Data Mining, Diabetes, T1D, AID, iCGM, Clustering, Regression, Decision-making, Time series, Cybersecurity, Methodology.*

## **1. INTRODUCTION**

The number of diabetic people in the world has increased over the years due to lack of physical exercise, poor diet habits, stress, and genetic factors. It is estimated that the population of patients with diabetes will double in 25 years. The expected number of patients by 2034 will be approximately 44 million individuals, and the costs related to treating the complications of the disease are US\$700 billion yearly. Thus, this paper seeks to bring intelligence to improve blood glucose control, using machine learning, reducing the ills of the disease and the costs related to disease treatments [24] [65].

To control diabetics, studies have been developed so that it is possible to administer medications with greater precision to patients. However, there is an excess of data being generated by multiple devices, based on the two main categories:

Blood glucose sensors - integrated continuous glucose monitoring (iCGM), such as the Freestyle Libre [8] [59] [14] and Dexcom [39].

Insulin pumps - automated insulin dosing systems (AID), such as the Tandem Diabetes Care t:Slim X2 insulin pump and the Ominipod.

According to Gordon Moore's observation [40], soon, several more wearables and IoTs, with even more technological capabilities should be available, further increasing the availability of data related to glucose control in patients with diabetes.

As the data volume tends to increase, it is inferred that an interoperable [34] system for [53] integration, indexing and data analysis, can benefit patients and healthcare systems through the use of a decision-making engine [37], using machine learning techniques.

This scenario may reduce the time to market [10] of other players in the glucose control scenario, increasing competitiveness and benefiting scientific production, in addition to reducing the costs of treating patients with acute diabetes.

Obviously, the provision of an artificial intelligence service to support the cloud decision-making algorithm [29], increases the attack surface for potential offenders [38]. Therefore, it is essential that advanced systems for cyberattacks early detection [41] are used to monitoring the intelligence service engine.

The Federal District Government, capital of Brazil, has a free iCGM program, via the Brazil's Unified Health System (SUS) [62], providing measurement sensors for patients with HbA1c up to 7 [64]. However, to get access to the sensors, the patient must fill out a document and then deliver it to a health centre in their region. Unfortunately, Brazil still has a great deficiency in the digital [36] and social [52] inclusion in a large part of the population. This fact makes it difficult to complete the aforementioned document, which can cause problems in accessing the blood glucose measuring device.

In this research, we used a project called AvantData for Good - AVD4G, by the Brazilian company Avantsec [5], where the goal is to collect data on diabetes control from FreeStyle Libre users [55], and the automatic completion of the report to ease the availability of the device to the disadvantaged population. The project also has, as a secondary goal, the storage, and processing of blood glucose data to support physicians and patients in the decision-making process of blood glucose control, respecting the General Data Protection Regulation [22] and Exchange of Supplementary Health Information (TISS) [20].

With the use and dissemination of AvantData for Good project [3] through the AvantAPI [4], it is expected to obtain a significant volume of data from diabetes patients, for clustering models algorithm [2], in order to determine possible groups of individuals and their relationships with factors influencing as: age, gender, race, volume of physical activity and eating habits of the population with diabetes.

## **2. BACKGROUND AND RELATED WORKS**

According to the International Diabetes Federation (IDF) [26], the number of children and teenagers living with diabetes increases annually. In 2019, more than one million children and adolescents had type 1 diabetes (T1D) and 86,000 children (increasing by 2 and 3% per year) are expected to develop T1D each year. There is currently no cure for T1D, but people with T1D with adequate daily insulin treatment, regular blood glucose monitoring, education and support can live healthy lives and delay or prevent many of the complications related to diabetes, therefore it is mandatory to keep glucose within the target range in patients with T1D. Although pharmacotherapeutic advances have improved in recent years, many patients are still not adept at

keeping their blood glucose levels within suggested limits. Blood Glucose Self-Monitoring (SMBG) helps improve blood glucose control and empowerment of people with diabetes; mainly useful for people with diabetes who are using insulin as it facilitates insulin delivery and detection of hypoglycemia.

In September 2016, the United States Food and Drug Administration (FDA) approved the Freestyle Libre Pro Glucose Monitoring System as a continuous glucose monitoring system (iCGM) for blind study in research. In September 2017 [19], the FDA approved the Freestyle Libre for personal use by patients, in which the sensor is disposable and applied to the back of the patient's arm, allowing its use for up to 14 days. Using this technology, physicians and patients can use a handheld device to download the blood glucose information stored in the sensor [8].

The "FreeStyle Libre" continuous glucose monitoring system [32] consists of a sensor and a handheld reader. A sensor unit is 3.5 cm in diameter and 0.5 cm thick. A thin filament is located in the centre underneath, which is coated with an adhesive. A handheld reader measures at a distance of 1 to 4 cm from the sensor. In addition to the glucose level, the device's touch-sensitive display indicates data curves as well as the trend of glucose values in arrow form. The sensor saves up to 8 hours of results. The handheld reader can be connected to a computer to create additional diagrams and analysis. A sensor measurement range from 2.2–27.8 mmol up to 1 (40–500 mg/dl). According to the manufacturer's specifications for use in humans, the sensor's lifetime is up to 14 days, before needing to be replaced [14].

Abbot [1] is the US company responsible for FreeStyle Libre product. The manufacturer makes a website available to its users [33] with the ability to collect and process information from (iCGM). Furthermore, the sensor works as an Analog Digital Converter (ADC)[47], therefore it allows storing and exported data.

To drug administration, it is possible to use insulin pump therapy, also known as automated insulin dosing systems (AID), which is an evolution application form, that has been shown to be highly effective in maintaining blood glucose levels and providing flexibility to patients' lives. It works by providing the patient with a continuous subcutaneous infusion of fast-acting insulin and allows the patient to administer bolus throughout the day, for feeding and correction of elevated glucose levels.

(AID) use is approved in patients with type 1 and selected patients with type 2 diabetes; however, it is important to select the correct patients for pump therapy. Insulin pump technology continues to evolve rapidly, and many options are now on the market, including those that are used in conjunction with continuous glucose monitoring (iCGM). However, (AID) is not suitable for all patients with diabetes who require the use of insulin [57].

Generally, AID is a treatment option for adults with type 1 diabetes who are motivated to improve glycemic control after a trial of multiple daily insulin injection (MDI) therapy and who can show the level of self-care required for adherence [23].

Artificial intelligence in medical applications using machine learning directly to the patient has been increasingly used to perform medical evaluation diagnoses [6]. Machine learning can make use of previously processed data to make objective and informed recommendations and can help ensure that decisions are made as assertively as possible, assisting healthcare professionals in decision-making.

The Algorithmic decision-making systems (ADMS) represent a holistic and interdisciplinary phenomenon and have some distinct interpretations [35]. Computer scientists focus on technical

aspects of algorithms, implicitly privilege commercial interests, and celebrate the power of the artifact, but there is some research to suggest that algorithmic systems create economic benefits that substantially exceed possible ethical concerns and implications [27, 48]. The other interpretation concerns sociologists and attorneys who are pessimistic and defend the prohibition of most ADMS because they are unfair, having unethical or inappropriate decisions can be triggered in society [54, 61].

Hence, with developing a pancreas that makes use of artificial intelligence, the development of an ADMS is intrinsic to decide from collected data, that is from the data collected from the iCGM, the decision-making system decision can trigger the AID to administer the amount of insulin needed on a historical basis.

### 3. METHODOLOGY FOR AN ARTIFICIAL INTELLIGENCE PANCREAS

To aid in the process of collecting glucose level data from patients, FreeStyle Libre was created as a less invasive technique that avoids the need for the patient to take a glucose measurement with multiple daily bites. After analysing the data, administration of the required daily insulin can be carried out after analysis by a physician using the insulin pump.

However, according to the FDA report [18], the FreeStyle Libre device may incorrectly indicate hypoglycemia, as after clinical studies it was verified that the device indicated 40% of the time that the user's sensor glucose values were at or below 60 mg/dL, where the user's glucose values were actually in the range of 81 to 160 mg/dL. For this reason, it is recommended that FreeStyle Libre glucose monitoring be based on trends and patterns analysed over time, not isolated data.

The scenario proposed in this paper for the control of diabetes in T1D patients is to use data mining and machine learning techniques to collect data from patients and, through clustering, provide enough information for decision-making algorithms with the use of artificial intelligence provides physicians with accurate information on administering the optimal amount of insulin according to the patient's identified glucose level. To achieve this goal, we suggest using the Methodology for an Artificial Intelligence Pancreas (MAIP) presented in Figure 1.

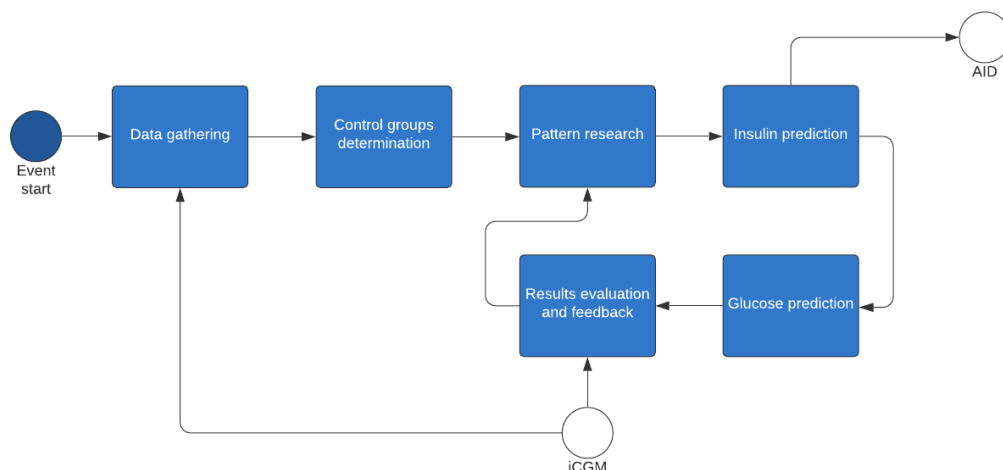


Figure 1. Methodology for an Artificial Intelligence Pancreas (MAIP)

The flow shown in Figure 1 is a continuous 6-step methodological process: Data collection, Determination of control groups, pattern search, insulin administration, blood glucose prediction, analysis of results and feedback.

### 3.1. Data gathering

Objective: Data receipt in AvantData for Good for dataset creation.

Processing: The website of Abbot, the company that created FreeStyle Libre, provides a converter for text spreadsheet, where all the readings sent can be consolidated for possible use by the user.

Expected outcome: It is suggested to export the data from the ADC sensor to a text file to load this information into AvantData for Good [4], which will be responsible for processing, filtering and indexing the data to create the dataset [30].

In the collection phase, there will be a short form that must be filled in by the patient with personal information. Data that allow the detection of whom the patient is will not be collected, that is, the objective is to have the minimum and necessary parameters for the creation of intelligence engines, but the individual's identity will be preserved and safeguarded.

### 3.2. Control groups determination

Objective: Detection of possible groups of patients with common characteristics (e.g., age, sex, cardiorespiratory capacity, physical activity, eating habits, among others).

Processing: Creation of a clustering model [63] in which the dataset must be prepared so that it can be processed. There are several pre-processing techniques available [25].

Dataset processing must be done using linear algebra [16], with mathematical operations between matrices and vectors. Thus, in order for the dataset to be properly processed, the categorical variables must be worked on to become their numerical equivalents [66].

After processing the dataset, it will be necessary to adjust the mathematical scales through feature scaling [51]. The use of min-max normalization is sufficient for this model, as the technique will allow the mitigation of possible mathematical dominance of a variable in relation to others in the model.

As we are dealing with an unsupervised model for this phase, it is not yet necessary to use the y vector. In this way, the dataset is ready for processing.

After pre-processing the data, the choice of classification models and the evaluation of the efficiency of each model begins, as well as the number of clusters that will be made available for the next phase. Each cluster will represent a group of individuals that will be studied in the control phase.

Due to the high accuracy demonstrated in other experiments [11], the K-Mean and SVM algorithms can be used to process the data in the clusters.

Once the model has been processed, the stage of adjusting the number of ideal clusters to define the control groups begins, as well as their related characteristics of associated factors determined in the "Data Collection" phase.

Expected outcome: Control groups correlated to factors associated with diabetes.

### 3.3. Pattern research

Objective: Establish a function of the relationship of the independent variables with the estimated probability for ideal insulin dosage [42].

Processing: Apply regression models to the control groups, using a physician-assisted insulin delivery process and with control of the glycemic response through the FreeStyle Libre of individuals belonging to the control groups [60, 9].

In this step, a mathematical function must be identified that will represent a method of estimating the volume of insulin for the patient. For this, the system will use non-linear regression algorithms, due to its satisfactory ability to reduce the risk of adverse situations [43].

Expected outcome: Analysis of data previously determined in the previous phase and creation of an insulin administration mechanism correlated with the individual's metabolic response and the expected blood glucose after a certain period of time. [21].

For each control group, it is suggested the administration of the medicine, via (AID), as prescribed by the physician and the sensing of the blood glucose level response via (iCGM). Thus, a new dataset will be created and can be represented by an X matrix, with the following independent variables: Control group, Amount of Insulin administered, Time of medicine administration, Current Insulin Level.

The y vector with the expected result of insulin in the next few minutes will serve as a basis for feedback to the system in later phases, as well as the measurement of performance in relation to the suggested dose [28].

Therefore, the objective of this step is to create models with high capacity to estimate the ideal insulin dose according to dataset data. This model will be evaluated and compared with the time series data presented in the next phases.

### 3.4. Insulin Prediction

Objective: Create a model that protects the patient against incorrect insulin administration. Obviously, insulin administration must be supervised by endocrinologists to ensure patient safety. At this stage, the objective is to determine a model to support decision-making, in order to protect the patient against any anomaly in the prediction stage.

Processing: With the result of the previous step, create a decision support algorithm [7] with the estimated amount of insulin for each individual [44], respecting their personal characteristics and receiving feedback in the closed-loop [56] of the FreeStyle Libre sensor (iCGM).

Based on the prediction of the amount of insulin administered via (AID) and the result obtained in several time intervals (iCGM), it is possible to obtain the creation of a new model where the blood glucose of a time interval (iCGM) will be analysed, related to the micro-dose of insulin (AID) applied in the immediately previous interval, that is, the blood glucose obtained (iCGM) after a certain period of time after the application of insulin via (AID).

With this information, the model must generate a classifier, which will be compared the estimated blood glucose, derived from the regression model, related to blood glucose achieved,

data obtained via (iCGM). The model will sort in a boolean way whether the administration has reached an expected level or not.

Thus, the system begins to have the fundamental building blocks for decision-making on whether to continue with the administration of the next micro-dose of insulin via (AID).

Expected outcome: The system must have the decision-making capacity to apply or not the next insulin micro-dose, and each insulin administration must be reported to the system, as well as the decision taken at that moment. Thus, the intelligence engine has enough data to refine the models and necessary improvements in order to increase the accuracy and mitigate the risks of inappropriately injecting rapid insulin into the patient's body.

### **3.5. Glucose Prediction**

Objective: The objective of this step is to create an estimate of blood glucose, distributed in time intervals in the future. In other words, predict how much blood glucose will be in the future, if this insulin is applied at the expected initial time.

Processing: Use a time series mechanism to predict the patient's blood glucose after insulin administration to provide feedback to the machine learning system and check whether the dose applied in the previous phase reached the expected result. Send feedback to the system with the expected blood glucose and the result achieved, measured via (iCGM).

With the proposed methodology, the effectiveness of the model must always be revised. To ensure that there is an efficient review process, another machine learning tool, based on time series, should be used to predict blood glucose in a short period of time [50].

Expected outcome: The time series will be based on the measured samples of the (iCGM), as well as the administration of rapid insulin (AID) and the date/time of release of the homonym. In other words, we are talking about a multivariate time series system for glucose forecasting [58]. This supports the main objective of the project to provide an intelligence engine that allows to simulate an artificial pancreas [12].

### **3.6. Results evaluation and Feedback**

Objective: The objective is to compare what was estimated with the individual's response and refine the model, seeking greater assertiveness in relation to the current medicine administration process.

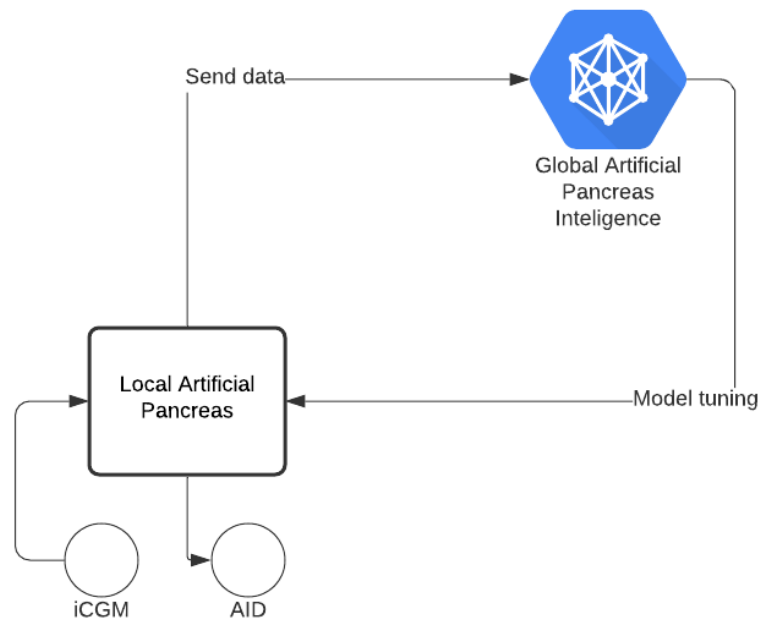


Figure 2. Collaborative Artificial Pancreas Intelligence

Processing: This cycle must be executed indefinitely, always looking for an optimal relationship for each patient.

Model evaluation techniques must be used and compared with the measured results to adjust the parameters in each of the phases.

Obviously, we will need special attention to avoid overfitting or under fitting. In other words, addicted models are unproductive and will not lead to adequate therapy.

Expected outcome: At the end of each step, the results must be collected and sent to the data collection step or for debugging the models, according to each case. Maintaining historical data over a period of time of at least 5 years is highly recommended as that way we can continually review the analyzed data, its benefits to patients and assess the overall performance of the system over time.

## 4. RESULTS

We conducted research to create a dataset, with more than 48 thousand blood glucose readings, as well as their statistical data were computed, via AvantAPI, for the first phase of the methodology, as represented in the table below:

Due to the high cost of the (AID) we are still unable to proceed with the other phases of the project for this paper. As the focus of the study is the methodology for feeding the model in closed-loop, it is not possible to continue detailing the results without the data from the (AID).



Table 1. Blood glucose reading of a patient

| Data                         | Max    | Min   | Mean   | Standard Deviation |
|------------------------------|--------|-------|--------|--------------------|
| Historic Glucose mg/dL       | 379.00 | 40.00 | 132.29 | 47.84              |
| Scan Glucose mg/dL           | 382.00 | 40.00 | 127.13 | 50.03              |
| Rapid-Acting Insulin (units) | 14.00  | 1.00  | 5.72   | 2.34               |
| Long-Acting Insulin (units)  | 50.00  | 3.00  | 43.72  | 4.49               |
| Total                        |        |       |        | 48,642             |

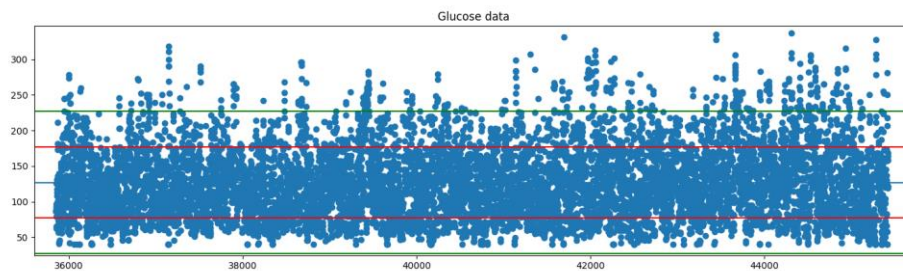


Figure 3. Glucose Sample Data

## 5. DISCUSSION

The proposed MAIP methodology seeks to support the work to create a cybernetic pancreas. In the same way that we have pacemakers, in the near future, we may have digital pancreas to support T1D diabetes therapy.

The scope of this work is limited to supporting cyber intelligence for the best performance of insulin delivery to the individual. This intelligence would be a common asset, in an open source, interoperable platform, so that any company or individual can use the benefits of a worldwide collaborative system to simulate the processing of carbohydrates in the human body.

For this infrastructure to be available on a global scale, 5G technology brings tools that were previously impractical. However, in the area of technology, there are several cyber threats, for example: ransomware, denial of service attacks and malware in general.

For the success of this project, it is necessary that users participate as collaborators, and for this they must be aware that the data collected will be used exclusively for research and their names and personal information will not be stored. It is important to emphasize the aspects related to the protection of privacy and possible cyberattacks. Therefore, it is essential to mention that the General Data Protection Regulation of each country must be strictly respected for the use of their information.

Regarding information security, the complete anonymization [49] of the data is adequate, however, for the clustering algorithm to be useful in the future, it is necessary to collect additional data and habits to analyze the correlation with the levels of each person's glucose.

During the process of determining the control groups, attention must be paid to the categorization of variables, as it can generate a problem of mathematical preponderance between a class in relation to others, due to the reason that each class becomes a number, and numbers represent

quantities that are mathematically comparable. But this is not the reality in the case of categories. For example: One race is not bigger or smaller than another, they are simply different. So, numbering the races would bring up the problem that one number is larger than the other, which is obviously not the case.

Thus, the ideal way would be to first transform each of the independent variables into a number and then each of the categories become other independent variables, represented in a boolean data. However, after this process, the dimensionality of the model will be increased.

Furthermore, it is necessary to monitor the dimensions of the model and adjust it, so that it is not too computationally burdensome, as the computational cost increases exponentially for each new dimension [15]. In theory, data sanitization should be restricted to disposing corrupted data. The data from the ADC are just about blood glucose and time, and this information will always be available. The other data related to associated factors comes from the user's registration, and the filling in of all information must be controlled in the data collection system.

Regarding the "Search for patterns" phase, it is noted that the vector  $y$  should initially be related to the expected response time after the administration of rapid insulin, however, as we have a continuous monitoring system (iCGM), the possibility of creating several  $y$  vectors with a regression model for each reading after the application of rapid insulin, where each one is related to the probabilistic value of the expected blood glucose in relation to the time of medicine administration. This process occurs due to the large mass of data measuring the individual's metabolic response every 5 minutes. Even with the ability to measure every minute [45] the manufacturer suggests a reading at a minimum interval of 5 minutes. Thus, this was the time interval chosen in this work.

The advantage of this process is that insulin can be administered continuously in micro-doses [13] and with monitoring in a short period of time. In this way, insulin administration becomes safer, as the system is continuously fed back and the decision-making mechanism has the necessary elements for safe administration, as the closed-loop delivers the necessary feedback to increase the level of consciousness situational assessment of the individual's metabolic profile.

This would be an ideal mechanism to simulate the functioning of the pancreas, due to the continuous measurements and responses with the release of insulin into the bloodstream. Obviously, with the caveat that insulin administration in this case is not made directly into the bloodstream [31].

The use of multiple dimensions in time series is widely used in other forecasting domains, such as weather forecasting. The advantage of the proposed methodology is that the probability of not having difficulties related to missing values is very high. This stems from a refined data acquisition process in the "Data Collection" phase of the methodology.

Another fundamentally important factor in predicting glucose after insulin release is that we will have feedback material to support decision-making and to improve the regression model. We will be able to study the effect of each administration, comparing the expected glucose with the value reached in the therapy of everyone.

## **6. LIMITATIONS AND THREATS TO VALIDITY**

Like any empirical study, this paper also has limitations and threats to its validity. As the study is a methodological proposal, a practical validation has not yet been carried out, collecting data from diabetic patients.

Due to the need to obtain numerous people during the process of implementing the proposed methodology, and to collect data from these patients to feed back to the system, it is necessary that patients are willing to provide the data, as the success of this project depends on primarily from the vast mass of data and assessments. In addition to glucose and insulin data, it is necessary to collect additional data such as: age, sex, volume of physical activity, if you are a user of other continuous-use medications, among other relevant data.

In order for this platform to be used on a global scale, an in-depth study of the best cyber defense practices is necessary in order to guarantee the availability, integrity and confidentiality of the collaborative effort. It would be useless for this intelligence to be available and the (AIDs) not to have adequate information on the measurement of insulin application.

In the same way, cyberterrorism attacks are another global threat and even attacks on patients' lives can occur, if security criteria are not very well worked out. This fact, in itself, obviously cannot be a demotivator for the research to continue. A joint effort with information security and cloud infrastructure companies is fully capable of enabling this technology to help improve the quality of life of T1D patients.

There are several possibilities to circumvent threats and make the system resilient and highly reliable. Such as blockchain to guarantee each of the transactions performed by the engine, as well as redundancies in (AID) and (iCGM). If we look more closely at Figure 3, we are talking about an artificial pancreas location, that is, even with the unavailability or integrity compromised, we would still have backups available with each of the users.

In summary, this material is not intended to exhaust the cybersecurity mechanisms to guarantee the system's functioning, but in-depth studies on the subject are of vital importance for the viability of the intelligence engine.

## **7. FUTURE WORKS**

For the correct definition of the variables to be used in the dataset, a complementary study is necessary, where the associated factors in the control of diabetes [17] must be analyzed. During data processing in clusters, the K-Mean and SVM algorithms can be used, however, the evaluation of the model requires further studies to ensure that the accuracy rate will be satisfactory.

The use of this methodology provides the structuring of a rich knowledge base, as future work can be carried out taking into account other associated factors, such as other syndromes or comorbidities. With a rich and cross-sectional basis for multivector analysis, we can think of asking new questions such as: What is the change in insulin therapy for patients affected with COVID-19? It is expected that with these data, this and several other responses will be answered more quickly, as we will have a historical memory of the individuals' metabolic response, as well as the effect caused by small nuances in the treatment.

The result of using the proposed methodology must be deeply validated with several endocrinologists to prevent the model from being wrongly trained and causing harm to the patient.

## 8. CONCLUSIONS

Studies indicate a significant increase in the quality of life of patients monitored via (iCGM), reaching a decrease in HbA1c of 0.3%. In addition to the 4-fold reduction in visits coded for diabetic ketoacidosis (DKA) in Emergency rooms. However, the control of insulin infusion via (AID) in relation to (MDI) does not represent a significant difference in hospitalization costs, as there is no gain in the use of (AID).

Thus, it is expected that with the model proposed in this work, assisted by artificial intelligence, for insulin administration, it will provide a gain in glycemic control and, consequently, an improvement in the patient's quality of life, as well as in the costs related to the hospitalization of patients from T1D [46].

## REFERENCES

- [1] Abbott: Abbott (November, 2021), <https://www.abbott.com/>, accessed in 11/26/2021
- [2] Alexandre, K., Vallet, F., Peytremann-Bridevaux, I., Desrichard, O.: Identification of diabetes self-management profiles in adults: A cluster analysis using selected self-reported outcomes. *Plos one* 16(1), e0245721 (2021)
- [3] Avantdata: Plataforma de análise, correlacionamento e gestão de dados em redes corporativas (December, 2018), <https://www.avantdata.com.br>, accessed in 11/25/2021
- [4] Avantdata: Documentação das chamadas ao novo avantapi (December 2020), <https://avantapi.avantsec.com.br/>, accessed in 11/25/2021
- [5] Avantsec: Avantsec - inovação em segurança da informação (December 2017), <https://www.avantsec.com.br/>, accessed in 11/25/2021
- [6] Babic, B., Gerke, S., Evgeniou, T., Cohen, I.G.: Direct-to-consumer medical machine learning and artificial intelligence applications. *Nature Machine Intelligence* 3(4), 283–287 (2021)
- [7] Bhargav, S., Kaushik, S., Dutt, V., et al.: A combination of decision trees with machine learning ensembles for blood glucose level predictions. In: *Proceedings of International Conference on Data Science and Applications*. pp. 533–548. Springer (2022)
- [8] Blum, A.: Freestyle libre glucose monitoring system. *Clinical Diabetes* 36(2), 203–204 (2018)
- [9] Borle, N.C., Ryan, E.A., Greiner, R.: The challenge of predicting blood glucose concentration changes in patients with type 1 diabetes. *Health Informatics Journal* 27(1), 1460458220977584 (2021)
- [10] Charney, C.: Time to market: reducing product lead time. *Society of Manufacturing Engineers* (1991)
- [11] Chauhan, T., Rawat, S., Malik, S., Singh, P.: Supervised and unsupervised machine learning based review on diabetes care. In: *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*. vol. 1, pp. 581–585. IEEE (2021)
- [12] Clarke, W.L.: The artificial pancreas: how close are we to closing the loop. *Pediatr Endocrinol Rev* 4, 314–316 (2007)
- [13] De Bock, M., Dart, J., Roy, A., Davey, R., Soon, W., Berthold, C., Retterath, A., Grosman, B., Kurtz, N., Davis, E., et al.: Exploration of the performance of a hybrid closed loop insulin delivery algorithm that includes insulin delivery limits designed to protect against hypoglycemia. *Journal of diabetes science and technology* 11(1), 68–73 (2017)
- [14] Deiting, V., Mischke, R.: Use of the “freestyle libre” glucose monitoring system in diabetic cats. *Research in veterinary science* 135, 253–259 (2021)
- [15] Donoho, D.L., et al.: High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture 1(2000)*, 32 (2000)
- [16] Elgohary, A., Boehm, M., Haas, P.J., Reiss, F.R., Reinwald, B.: Compressed linear algebra for large-scale machine learning. *Proceedings of the VLDB Endowment* 9(12), 960–971 (2016)
- [17] Faria, H.T.G., Rodrigues, F.F.L., Zanetti, M.L., Araújo, M.F.M.d., Damasceno, M.M.C.: Fatores associados à adesão ao tratamento de pacientes com diabetes mellitus. *Acta Paulista de Enfermagem* 26, 231–237 (2013)
- [18] Food, (FDA), D.A.: Summary of safety and effectiveness data (ssed) (September 2016), [https://www.accessdata.fda.gov/cdrh\\_docs/pdf15/p150021b.pdf](https://www.accessdata.fda.gov/cdrh_docs/pdf15/p150021b.pdf), accessed in 11/26/2021

- [19] Food, (FDA), D.A.: Fda approves first continuous glucose monitoring system for adults not requiring blood sample calibration (September 2017), <https://www.fda.gov/news-events/press-announcements/fda-approves-first-continuous-glucose-monitoring-system-adults-not-requiring-blood-sample>, accessed in 11/26/2021
- [20] Galvão, M.C.B.: Classificações, terminologias e ontologias no campo da saúde. *Asklepion: Informação em Saúde* 1(2), 41–54 (2021)
- [21] Hart, D.: Artificial intelligence & machine learning for effective management of blood glucose levels in patients with type 1 diabetes. Tech. rep., EasyChair (2021)
- [22] Hawryliszyn, L.O., Coelho, N.G.S.C., Barja, P.R.: Lei geral de proteção de dados (lgpd): O desafio de sua implantação para a saúde. *Revista Univap* 27(54) (2021)
- [23] Heinemann, L., Fleming, G.A., Petrie, J.R., Holl, R.W., Bergenstal, R.M., Peters, A.L.: Insulin pump risks and benefits: a clinical appraisal of pump safety standards, adverse event reporting and research needs. a joint statement of the european association for the study of diabetes and the american diabetes association diabetes technology working group. *Diabetologia* 58(5), 862–870 (2015)
- [24] Huang, E.S., Basu, A., O’grady, M., Capretta, J.C.: Projecting the future diabetes population size and related costs for the us. *Diabetes care* 32(12), 2225–2229 (2009)
- [25] Huang, J., Li, Y.F., Xie, M.: An empirical analysis of data preprocessing for machine learning-based software cost estimation. *Information and software Technology* 67, 108–127 (2015)
- [26] (IDF), I.D.F.: International diabetes federation (idf) (November 1950), <https://idf.org/>, accessed in 27/11/2021
- [27] Jarrahi, M.H., Sutherland, W.: Algorithmic management and algorithmic competencies: Understanding and appropriating algorithms in gig work. In: Taylor, N.G., Christian-Lamb, C., Martin, M.H., Nardi, B.A. (eds.) *Information in Contemporary Society - 14th International Conference, iConference 2019, Washington, DC, USA, March 31 - April 3, 2019, Proceedings. Lecture Notes in Computer Science*, vol. 11420, pp. 578–589. Springer (2019). [https://doi.org/10.1007/978-3-030-15742-5\\_55](https://doi.org/10.1007/978-3-030-15742-5_55), [https://doi.org/10.1007/978-3-030-15742-5\\_55](https://doi.org/10.1007/978-3-030-15742-5_55)
- [28] Kølendorf, K., Bojsen, J., Deckert, T.: Clinical factors influencing the absorption of 125i-nph insulin in diabetic patients. *Hormone and Metabolic Research* 15(06), 274–278 (1983)
- [29] Kuyoro, S., Ibikunle, F., Awodele, O.: Cloud computing security issues and challenges. *International Journal of Computer Networks (IJCN)* 3(5), 247–255 (2011)
- [30] Lau, A., Passerat-Palmbach, J.: Statistical privacy guarantees of machine learning preprocessing techniques. arXiv preprint arXiv:2109.02496 (2021)
- [31] Lewis, D.M.: Do-it-yourself artificial pancreas system and the openaps movement. *Endocrinology and Metabolism Clinics* 49(1), 203–213 (2020)
- [32] Libre, F.: Freestyle libre (December 2017), <https://www.freestylelibre.de/>, accessed in 11/26/2021
- [33] Libre, F.: Freestyle libre (November 2021), <https://www.libreview.com/>, accessed in 11/26/2021
- [34] Lopez, D.M., Blobel, B.G.: A development framework for semantically interoperable health information systems. *International journal of medical informatics* 78(2), 83–103 (2009)
- [35] Marabelli, M., Newell, S., Handunge, V.: The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *The Journal of Strategic Information Systems* 30(3), 101683 (2021)
- [36] Mattos, F.A.M.d., Chagas, G.J.d.N.: Desafios para a inclusão digital no brasil. *Perspectivas em Ciência da Informação* 13, 67–94 (2008)
- [37] Meyer, G., Adomavicius, G., Johnson, P.E., Elidrissi, M., Rush, W.A., Sperl-Hillen, J.M., O’Connor, P.J.: A machine learning approach to improving dynamic decision making. *Information Systems Research* 25(2), 239–263 (2014)
- [38] Milan, S., Hintz, A.: Networked collective action and the institutionalized policy debate: bringing cyberactivism to the policy arena? *Policy & Internet* 5(1), 7–26 (2013)
- [39] Monitoring, D.C.G.: Dexcom continuous glucose monitoring (November 2021), <https://www.dexcom.com/>, accessed in 27/11/2021
- [40] Moore, G.E., et al.: Progress in digital integrated electronics. In: *Electron devices meeting*. vol. 21, pp. 11–13. Washington, DC (1975)
- [41] Narayanan, S., Ganesan, A., Joshi, K., Oates, T., Joshi, A., Finin, T.: Cognitive techniques for early detection of cybersecurity events. arXiv preprint arXiv:1808.00116 (2018)

- [42] Nguyen, M., Jankovic, I., Kalesinskas, L., Baiocchi, M., Chen, J.H.: Machine learning for initial insulin estimation in hospitalized patients. *Journal of the American Medical Informatics Association* 28(10), 2212–2219 (2021)
- [43] Noaro, G., Cappon, G., Sparacino, G., Del Favero, S., Facchinetti, A.: Nonlinear machine learning models for insulin bolus estimation in type 1 diabetes therapy. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 5502–5505. IEEE (2020)
- [44] Obeidat, Y., Ammar, A.: A system for blood glucose monitoring and smart insulin prediction. *IEEE Sensors Journal* 21(12), 13895–13909 (2021)
- [45] Olafsdottir, A.F., Attvall, S., Sandgren, U., Dahlqvist, S., Pivodic, A., Skrtic, S., Theodorsson, E., Lind, M.: A clinical trial of the accuracy and treatment experience of the flash glucose monitor freestyle libre in adults with type 1 diabetes. *Diabetes technology & therapeutics* 19(3), 164–172 (2017)
- [46] Parkin, C.G., Graham, C., Smolskis, J.: Continuous glucose monitoring use in type 1 diabetes: longitudinal analysis demonstrates meaningful improvements in hba1c and reductions in health care utilization. *Journal of diabetes science and technology* 11(3), 522–528 (2017)
- [47] De la Paz, E., Barfidokht, A., Rios, S., Brown, C., Chao, E., Wang, J.: Extended noninvasive glucose monitoring in the interstitial fluid using an epidermal biosensing patch. *Analytical Chemistry* 93(37), 12767–12775 (2021)
- [48] Priami, C.: Algorithmic systems biology. *Commun. ACM* 52(5) 80–88 (2009) <https://doi.org/10.1145/1506409.1506427>, <https://doi.org/10.1145/1506409.1506427>
- [49] Ribeiro, S.L., Nakamura, E.T.: Privacy protection with pseudonymization and anonymization in a health iot system: results from ocariot. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE). pp. 904–908. IEEE (2019)
- [50] Rodríguez-Rodríguez, I., Chatzigiannakis, I., Rodríguez, J.V., Maranghi, M., Gentili, M., Zamora-Izquierdo, M. Á.: Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques. *Sensors* 19(20), 4482 (2019)
- [51] Rohini, M., Surendran, D.: Toward alzheimer’s disease classification through machine learning. *Soft Computing* 25(4), 2589–2597 (2021)
- [52] de Senne, F.J.N.: Desigualdades digitais e inclusão social: uma análise da trajetória do acesso e uso da internet em regiões metropolitanas brasileiras
- [53] Shapiro, J.S., Mostashari, F., Hripcsak, G., Soulikis, N., Kuperman, G.: Using health information exchange to improve public health. *American journal of public health* 101(4), 616–623 (2011)
- [54] Shin, D., Y.J.: Role of fairness, accountability and transparency algorithmic affordance. *Computat. Hum., Behav.* 98, (2019). <https://doi.org/10.1016/j.chb.2019.04.019>, <https://doi.org/10.1016/j.chb.2019.04.019>
- [55] Silk, A.D.: Diabetes device interoperability for improved diabetes management. *Journal of diabetes science and technology* 10(1), 175–177 (2016)
- [56] Song, S., Song, K., Xu, T., Zhou, W., Li, H., Liu, W.: The electronics design of real-time feedback control system in ktx. *IEEE Transactions on Nuclear Science* (2021)
- [57] Sora, N.D., Shashpal, F., Bond, E.A., Jenkins, A.J.: Insulin pumps: review of technological advancement in diabetes management. *The American journal of the medical sciences* 358(5), 326–331 (2019)
- [58] Tang, X., Yao, H., Sun, Y., Aggarwal, C., Mitra, P., Wang, S.: Joint modeling of local and global temporal dynamics for multivariate time series forecasting with missing values. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 34, pp. 5956–5963 (2020)
- [59] Tsoukas, M., Rutkowski, J., El-Fathi, A., Yale, J.F., Bernier-Twardy, S., Bossy, A., Pytka, E., Legault, L., Haidar, A.: Accuracy of freestyle libre in adults with type 1 diabetes: the effect of sensor age. *Diabetes technology & therapeutics* 22(3), 203–207 (2020)
- [60] Tucker, A.P., Erdman, A.G., Schreiner, P.J., Ma, S., Chow, L.S.: Examining sensor agreement in neural network blood glucose prediction. *Journal of Diabetes Science and Technology* p. 19322968211018246 (2021)
- [61] Veale, M., Kleek, M.V., Binns, R.: Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In: Mandryk, R.L., Hancock, M., Perry, M., Cox, A.L. (eds.) *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018*,

- Montreal, QC, Canada, April 21-26, 2018. p. 440. ACM (2018). <https://doi.org/10.1145/3173574.3174014>, <https://doi.org/10.1145/3173574.3174014>
- [62] Vencio, S., Caiado, A., Morgental, D., Bufaiçal, N., Carneiro, R., Vencio, R.C.: 21st brazilian diabetes society congress (2018)
- [63] Vimal, V.: Prediction of diabetes mellitus using machine learning algorithms. In-ternational Journal of Advanced Engineering Science and Information Technology 4(4) (2021)
- [64] Weykamp, C.: Hba1c: a review of analytical and clinical aspects. Annals of laboratory medicine 33(6), 393–400 (2013)
- [65] Wu, X., Guo, X., Zhang, Z.: The efficacy of mobile phone apps for lifestyle modifica-tion in diabetes: systematic review and meta-analysis. JMIR mHealth and uHealth 7(1), e12297 (2019)
- [66] Zelaya, C.V.G.: Towards explaining the effects of data preprocessing on machine learning. In: 2019 IEEE 35th international conference on data engineering (ICDE). pp. 2086–2090. IEEE (2019)

## AUTHORS

### Rodrigo Vilanova

Higher Education Graduation in Information Security Technology at ICESP on March 1, 2007.

Fluency in English. Native Portuguese. Knowledge of Spanish and Mandarin.

Creator of the AvantData product suite (<https://www.avantdata.com.br>).



### Anderson Jefferson Cerqueira

I am a PhD student in Computer, I hold a Master's Degree in Application Computer from University of Brasília (UnB), Brazil in 2019. MBA in Software Engineering (2009) and graduated in Systems Information (2007). I'm Professor of Computer Science Course since 2009. My research interests include Software Engineering, Government Technology, Cybernetics and Health Informatics.

