

## Research Article

# Application of Artificial Intelligence Techniques for the Estimation of Basal Insulin in Patients with Type I Diabetes

Guillermo Edinson Guzman Gómez <sup>1,2</sup>, Luis Eduardo Burbano Agredo,<sup>3</sup>  
Veline Martínez,<sup>2,3</sup> and Oscar Fernando Bedoya Leiva<sup>4</sup>

<sup>1</sup>Fundación Valle del Lili, Departamento de Endocrinología, Cali, Colombia

<sup>2</sup>Universidad Icesi, Facultad de Ciencias de la Salud, Cali, Colombia

<sup>3</sup>Universidad del Valle, Cali, Colombia

<sup>4</sup>Universidad del Valle, School of Computer Science and Systems Engineering, Cali, Colombia

Correspondence should be addressed to Guillermo Edinson Guzman Gómez; guillermoguzman@gmail.com

Received 16 June 2020; Revised 2 October 2020; Accepted 15 October 2020; Published 3 November 2020

Academic Editor: Peng Fei Shan

Copyright © 2020 Guillermo Edinson Guzman Gómez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Artificial intelligence techniques have been positioned in the resolution of problems in various areas of healthcare. Clinical decision support systems developed from this technology have optimized the healthcare of patients with chronic diseases through mobile applications. In this study, several models based on this methodology have been developed to calculate the basal insulin dose in patients with type I diabetes using subcutaneous insulin infusion pumps. *Methods.* A pilot experimental study was performed with data from 56 patients with type 1 diabetes who used insulin infusion pumps and underwent continuous glucose monitoring. Several models based on artificial intelligence techniques were developed to analyze glycemic patterns based on continuous glucose monitoring and clinical variables in order to estimate the basal insulin dose. We used neural networks (NNs), Bayesian networks (BNs), support vector machines (SVMs), and random forests (RF). We then evaluated the agreement between predicted and actual values using several statistical error measurements: mean absolute error (MAE), mean square error (MSE), root-mean-square error (RMSE), Pearson's correlation coefficient ( $R$ ), and determination coefficient ( $R^2$ ). *Results.* Twenty-four different models were obtained, one for each hour of the day, with each chosen technique. Correlation coefficients obtained with RF, SVMs, NNs, and BNs were 0.9999, 0.9921, 0.0303, and 0.7754, respectively. The error increased between 06:00 and 07:00 and between 13:00 and 17:00. *Conclusions.* The performance of the RF technique was excellent and got very close to the actual values. Intelligence techniques could be used to predict basal insulin dose. However, it is necessary to explore the validity of the results and select the target population. Models that allow for more accurate levels of prediction should be further explored.

## 1. Introduction

Diabetes mellitus (DM) is a chronic disease whose prevalence is on the rise and has been called the epidemic of the 21st century. DM has a great implication on morbidity and mortality. This disease represents high costs derived from the care of patients and the complications resulting from poor control of the disease [1]. Type 1 DM (T1DM) is characterized by low insulin production by the pancreas and the treatment is based on an exogenous replacement. Insufficient insulin in the body can lead to serious

complications, such as ketoacidosis and hyperosmolar status that can lead to a diabetic coma and death [2, 3].

Many factors can affect blood glucose levels such as exercise, viral diseases, stress, and hormonal effects, which are not related to periods of food intake or insulin bolus but might be present during the time that basal insulin must cover. These events directly impact blood glucose levels in T1DM patients, hence the importance of establishing a correct basal insulin dose from the beginning of the treatment while facilitating the adjustment of the dose corresponding to boluses and sensitivity. Currently, medical

guides propose insulin schedules based on the weight and age of the patient without taking into account other factors, such as those mentioned above [4]. Furthermore, very often, insulin adjustments end up being established following simple rules based on empiricism and the experience of the treating physician and are implemented late due to the difficulty the patient faces accessing specialized consultations.

In emerging countries like Colombia, access to health-care services is limited. In this regard, clinical decision support systems based on artificial intelligence techniques could be useful as a strategy to close the care gaps [5–9]. Currently, different models have been proposed to estimate the basal insulin dose using artificial intelligence algorithms and techniques [10–18]. These models have shown good results in improving the time in the range of glycemic control [10–13]. Consequently, in this study, we aimed to evaluate different artificial intelligence models calculating basal insulin doses and estimates which could be better.

## 2. Material and Methods

A retrospective cohort study was conducted analyzing data from adult patients over 18 years with a diagnosis of T1DM using Medtronic 640G and Paradigm Veo insulin infusion pumps coupled to continuous glucose monitoring using Enlite sensor and with an acceptable glycemic control, defined by a range of glycated hemoglobin (HbA1c) between 6% and 8% and sensor use more than 80% were included. Comorbidities were not considered to excluded patients. All were treated at the Valle del Lili Foundation, located in Cali, Colombia.

Variables such as age, sex, weight, height, body mass index, glycated hemoglobin, basal insulin dose, upper and lower bounds of the standard deviation of the insulin dose, total insulin dose, percentage of basal insulin use and the daily dose of basal insulin, interstitial glucose, and upper and lower bounds of the standard deviation of interstitial glucose measurements were included. Data were obtained from the medical records collected using Medtronic Carelink Pro Version 4.0c software with a two-week data downloaded and supplemented with the clinical histories of the selected patients. For data analysis, we excluded days where high variability in blood glucose levels was observed (presence of acute disease, use of temporary basal conditions, hyperglycemia due to poor carbohydrate count, etc.).

**2.1. Models for Estimating Basal Insulin Dose.** Four artificial intelligence techniques were explored to estimate the basal insulin dose for each of the 24 hours in a day: neural networks (NNs), Bayesian networks (BNs), random forests (RF), and support vector machines (SVMs). The techniques were selected based on their ease of implementation in the R programming language. The artificial intelligence techniques selected are known as machine learning strategies. They create an insulin dose prediction model based on information from patients who are part of a training set. The model is capable of predicting insulin dose values for

patients to different test set from the training set and therefore, could be generalized. The prediction model varies according to the machine learning technique. Machine learning strategies do not use a simple mathematical formula to predict a patient's dose. Instead, they train a model presenting values that correspond to insulin doses of patients in a training set. In this way, the model is able to learn the relationship between variables that represent each patient and insulin dose in order to predict the dose in patients from a test set different from the training set. For training, testing, and cross-validation of the models, we used the information of 56 patients, from whom a set of 640 records was obtained corresponding to the glycemic data for 15 days and for each of the 24 hours in a day.

**2.2. Model Based on Neural Networks.** Artificial neural networks are computational models analogous to biological neural networks present in the nervous system of living beings. Neural networks are made up of nodes or units connected through directed connections. A connection from unit “ $j$ ” to unit “ $i$ ” serves to propagate activation “ $a$ ” from “ $j$ ” to “ $i$ ”. Furthermore, each connection has an associated numerical weight “ $W_{j,i}$ ” which determines the strength and sign of the connection. Each unit “ $i$ ” first calculates a weighted sum of its inputs:

$$in_i = \sum_{j=0}^n W_{j,i} a_j. \quad (1)$$

Then an activation function  $g$  is applied to this sum to produce the output:

$$a_i = g(in_i) = g\left(\sum_{j=0}^n W_{j,i} a_j\right). \quad (2)$$

The model based on the NNs technique was obtained using the neural net library in R. The network has 13 input neurons that correspond to the 13 aforementioned variables [11, 12]. Different topologies with different configurations were used and a 13-7-3-2-1 network topology was selected, that is, 13 input neurons, corresponding to the variables mentioned above, and 7, 3, and 2 neurons in hidden layers. Finally, in the output layer, only one neuron called Dose is obtained. The heuristic used was  $(n + m)/2$ , where  $n$  corresponds to the number of inputs and  $m$  to the number of output neurons. For example, for the first hidden layer, we have  $(13 + 1)/2$ . For this reason, 7 is the number of neurons in the first hidden layer. For the following layers, empirical adjustments determined from the results were performed. The best results were obtained by implementing the stochastic average gradient (SAG) algorithm, which is based on the Grprop algorithm and also derives from the electropropagation algorithm, defining the learning rate with the smallest absolute derivative without backward movement. The topology used is presented in Figure 1.

**2.3. Model Based on Bayesian Networks.** A Bayesian network is a directed acyclic graph made up of nodes and edges,



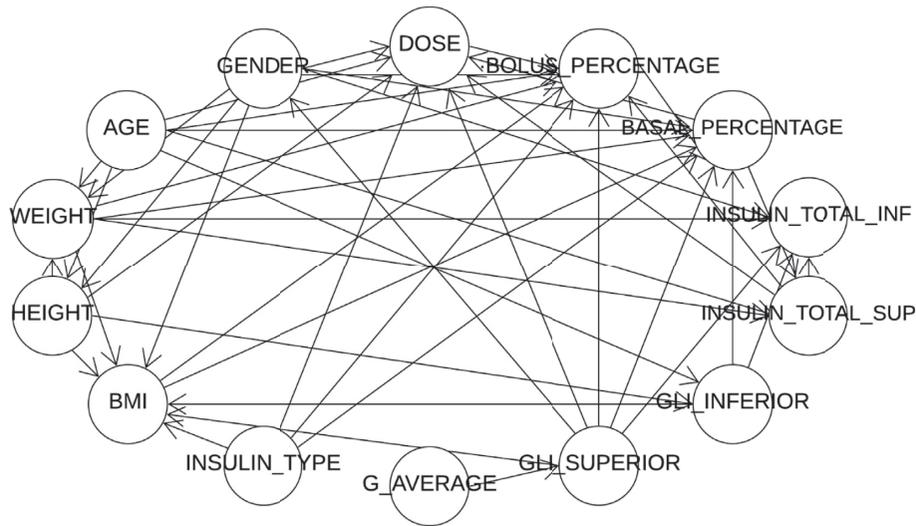


FIGURE 2: Dependencies in the Bayesian network model.

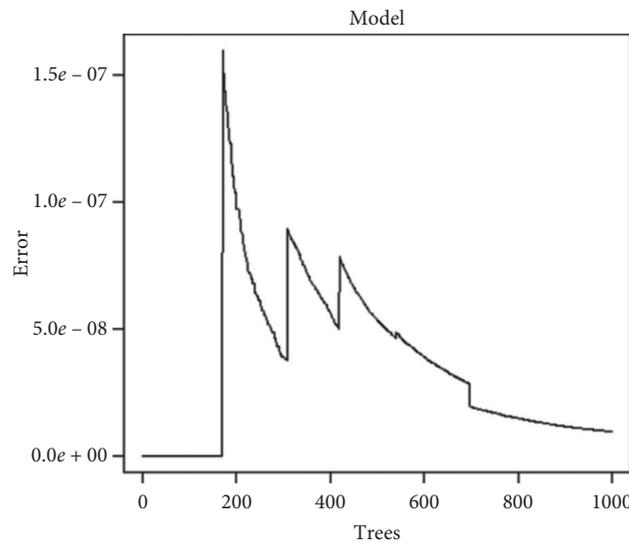


FIGURE 3: Representation of the prediction error using Random Forest.

### 3. Result Analysis

We evaluated the concordance of the four artificial intelligence techniques in the calculation of the basal insulin dose in relation to the dose established by the physician defined in the data download of the insulin pump. Metrics used to measure accuracy were as follows: mean absolute error (MAE), mean squared error (MSE), root-mean-square error (RMSE), Pearson's correlation coefficient ( $R$ ), and coefficient of determination ( $R^2$ ).

### 4. Results

Only 58 out of 480 patients had a regular follow-up at the institution from 2012 to 2018 and an acceptable glycemic control, i.e., a range of HbA1c between 6% and 8%. About 1456 records were stored for each patient. This data, a total of 81,536 records, were used to train the models (see Table 1).

The metrics used to measure the accuracy of the four techniques employed are shown using a color scale in Figure 4. The mean absolute error (MAE) of the four techniques (BNs, NNs, RF, and SVMs) is plotted on the X-axis and the 24 experiments (designed for each model as indicated in 2.4) are plotted on the Y-axis.

As observed in Figure 4(a), the BN technique reaches mean absolute errors (MAE) close to 1 which are represented in red. It indicates that the BN technique is not able to predict basal insulin doses specifically in certain hours such as 9 am, 2 pm, 3 pm, and between 8 pm and 11 pm. On the other hand, the RF and SVM techniques obtained mean absolute errors close to 0 which are presented in light green. Figure 4(a) allows seeing that RF and SVM techniques are more suitable for predicting basal insulin doses than the Bayesian networks and Neural networks. Figure 4(b) shows the color scale for the root-mean-square error (RMSE). In this case, similar results are obtained. However, the SVM

TABLE 1: Demographic data.

| Characteristics              | Average value | Standard deviation |
|------------------------------|---------------|--------------------|
| Age (years)                  | 39.81         | 11.56              |
| Women (%)                    | 77.4          | NA                 |
| Weight (kg)                  | 65.81         | 8.45               |
| Height (cm)                  | 162.96        | 8.73               |
| BMI (kg/m <sup>2</sup> )     | 24.77         | 2.61               |
| Total insulin dose (UI)      | 37.78         | 15.5               |
| Basal insulin (IU/hr)        | 0.908         | 0.43               |
| %basal insulin               | 51.23         | 10.12              |
| Interstitial glucose (mg/dL) | 138.19        | 12.99              |
| HbA1c (%)                    | 7.1           | 0.66               |

NA = not applicable.

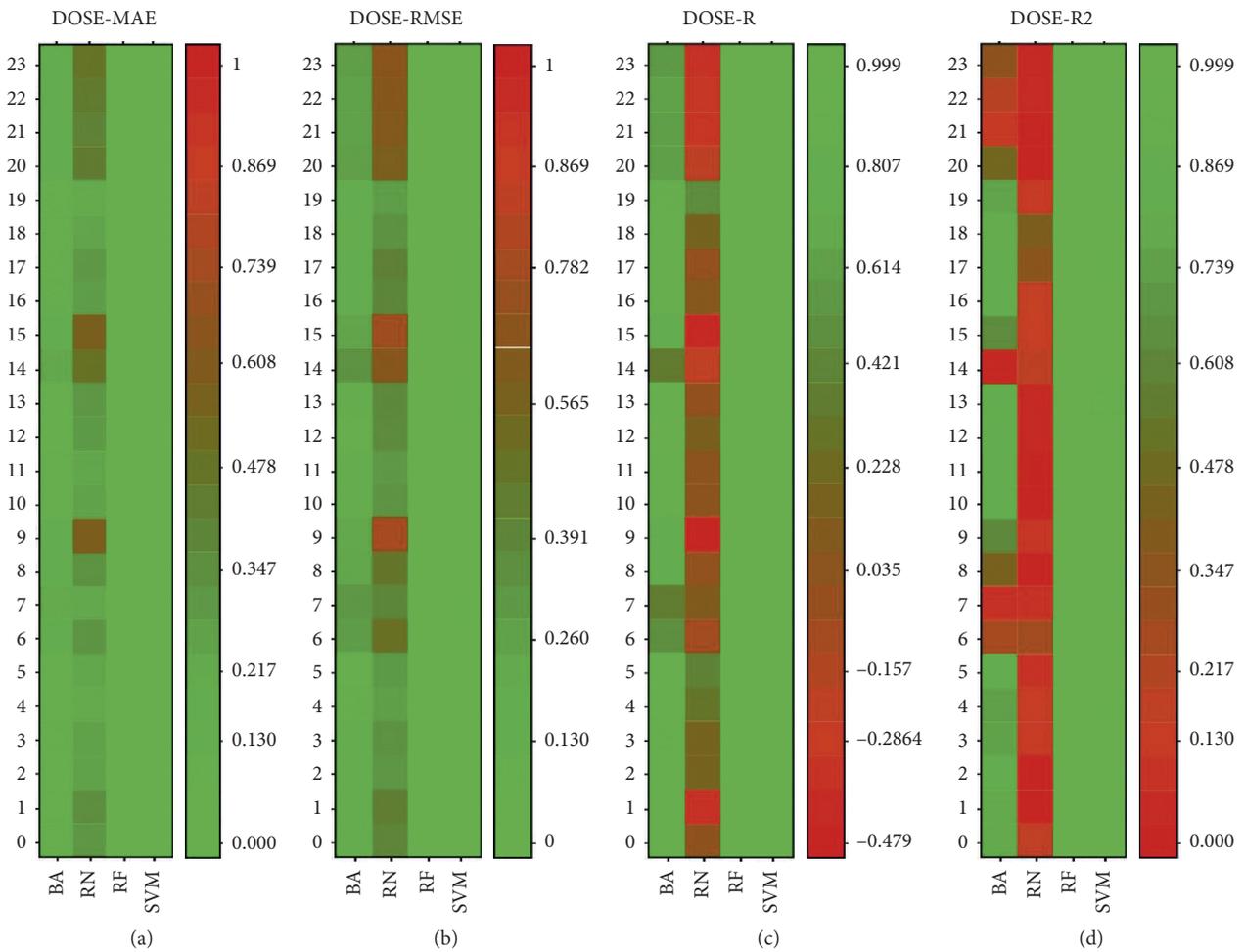


FIGURE 4: The metrics used to measure the accuracy of the four artificial intelligence techniques for the 24 subsets of data for dose calculation (a). Mean absolute error (b). Root-mean-square error (c). Pearson’s correlation coefficient (d). Coefficient of determination.

technique shows some values between 2 am and 7 am, and between 1 pm and 5 pm that have a dark green color which indicates that the RMSE values are not very close to 0. It should be observed that light green color is obtained in all hours for the RF technique. It means that Random Forest is better than BN, RN, and SVMs for predicting basal insulin doses. Figure 4(c) shows the color scale for the Pearson’s correlation coefficient (R). In this figure, the green color

represents a correlation close to 1.0. It occurs when the predicted and the actual values are highly correlated. A Person correlation coefficient of +1 is called a total positive linear correlation and it shows that a technique is suitable in a specific prediction problem. As observed, all values are light green in RF and SVM which indicates an excellent correlation between the predicted and actual values. On the other hand, BN and RN have some negative correlation

TABLE 2: Metrics used to measure the accuracy of experiments for dose calculation.

| Technique | MAE             | MSE              | RMSE           | R             | R <sup>2</sup> |
|-----------|-----------------|------------------|----------------|---------------|----------------|
| BNs       | 0.1199083333    | 0.03133645935    | 0.1770210704   | 0.7754166667  | 0.634873650    |
| NNs       | 0.3369750000    | 0.20192390070    | 0.4493594338   | -0.0303916666 | 0.067199665    |
| RF        | 0.0000657431064 | 0.00000037425226 | 0.000611761611 | 0.9999964346  | 0.999992869    |
| SVM       | 0.01282676086   | 0.001109456742   | 0.03330850856  | 0.9921356180  | 0.984373906    |

MAE: mean absolute error. MSE: mean squared error. RMSE: root-mean-square error. R: Pearson's correlation coefficient. R<sup>2</sup>: coefficient of determination. BNs: Bayesian networks. NNs: neural networks. RF: Random Forest. SVMs: support vector machines.

values that are presented in red. Finally, Figure 4(d) shows the color scale for the coefficient of determination (R<sup>2</sup>). The coefficient of determination measures how well a technique approximates the actual insulin doses. An R<sup>2</sup> of 1 indicates that a technique perfectly fits the data. As observed, Random Forest presents all R<sup>2</sup> values in light green. It means that RF is the best technique for predicting insulin doses. SVM has just a few hours in a darker green color. The BN and RN techniques have a poor coefficient of determination.

In the Table 2, we present a summary of the metrics used to measure the accuracy of the five experiments using the four techniques. The results obtained in the statistical error measurements show that the BNs and NNs techniques do not estimate the dose properly, while RF and SVMs had an excellent behavior getting very close to actual values with a MAE of 0.00006574310642 for RF and 0.01282676086 for SVMs. It was also evidenced that the error increases in certain periods of time, from 6 am to 7 am and between 13 pm to 17 pm. Pearson's correlation coefficient (R) with RF, SVMs, NNs, and BNs, were 0.9999, 0.9921, 0.0303, and 0.7754, respectively. As explained before, the values obtained by RF and SVM are considered a total positive linear correlation. It shows that RF and SVM are excellent choices when predicting insulin doses. The coefficient of determination (R<sup>2</sup>) of RF, SVMs, NNs, and BNs, was 0.9999, 0.9843, 0.0671, and 0.6348, respectively. The R<sup>2</sup> value obtained by the RF technique is very close to 1.0 and it means that Random Forest is highly accurate to approximate the actual insulin doses.

In total, 96 artificial intelligence models were designed. It means 24 different models for each technique used, considering that the pump model used in the study allows a maximum of 24 basal insulin doses to be configured. The 24 models provide a basal insulin dose for each hour. Although these doses are not necessarily different, according to the tests carried out, there may be several doses per hour that can constitute a single dose range. Figure 5 shows an estimate of 24 insulin doses calculated for 24 hours a day in a patient and how the dose estimated in some range of times were similar, for example, the estimated doses between 18:00-19:00 = 1,029 IU/hr. and 19:00-20:00 = 1,006 IU/hr.

## 5. Discussion

DM is a highly prevalent condition worldwide with many nuances and to control it, a multidisciplinary approach with patient participation is required. Therefore, it is important to create strategies which allow the empowerment of patients to achieve glucose level goals avoiding hyperglycemia or

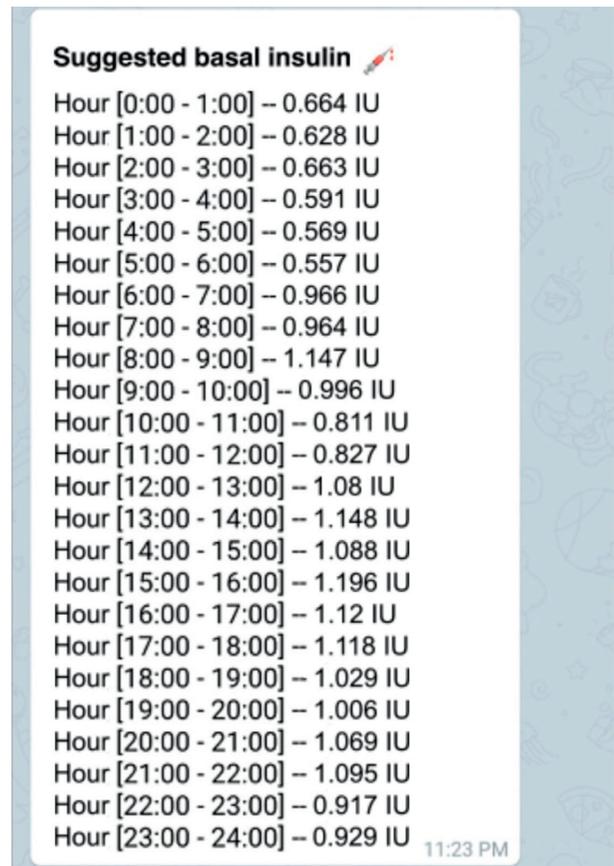


FIGURE 5: Example of basal dose insulin created by models.

hypoglycemia and their impact on morbidity and mortality. To this end, an adequate calculation of the basal insulin dose is an important strategy. Clinical decision support systems have been emerging as alternative methods in the care of patients with chronic diseases, such as diabetes. However, a comprehensive validation process is required [19, 20] for their application in patients and to achieve the proposed objectives.

Currently, different models have been proposed to predict the basal insulin dose using algorithms and artificial intelligence techniques [10–18]. Torrent-Fontbona proposes a system that calculates boluses with case-based reasoning and that in combination with Kalman filtering for baseline estimation reaches a time in the range of 83.87 + 1.35 [10, 11]. Meanwhile, Cappon et al. [12] use a NN to calculate insulin bolus with a reduction in the blood glucose risk index of 0.37, 0.23, and 0.20 versus the standard methods Scheiner, Pettus, and Edelman,

respectively. In the study carried out by Nimri et al. [13] with 25 patients using fuzzy logic, most patients had numerical differences between  $-0.5$  and  $6$  units with respect to the total daily dose of insulin prescribed by the medical specialist. However, there were cases where the difference was greater. So far, the results have been good, for example, in one of these cases, it was possible to go from an average glycemia of  $259.43 \pm 23.42$  mg/dL to  $156.73 \pm 14.41$  mg/dL.

In our case, of all the models obtained, The RF technique was found to be the most suitable for estimating the basal insulin dose and the dose estimated in some range of times was similar this according to the configuration that allows the insulin pump model 640G. Further, the error was greater in certain periods of time, between 6 am and 7 am and between 1 pm and 5 pm. This may be because they are the periods with the highest glycemic variability (GV), an integral component of glucose homeostasis [21]. Diet plays an important role in the variation of glycemia [22–24]. This concurs with the periods when calculating the basal insulin dose was less successful and coincided with breakfast in the morning and lunch in the afternoon. It will be important to join strategies in the future to optimize the calculation of ingested carbohydrates. Additionally, there might be other factors related to GV that should be taken into account to improve the model, such as stress, exercise, comorbidities, and the use of other medications.

Despite the fact that the number of patients seems small, records used for model training are much larger and more representative of the patients. A total of 1456 records were stored for each patient. Thus, 81,536 records were used to train models. This training stage allows the models to have the ability to generalize and be applied for basal insulin dose prediction to patients who did not participate in the training while maintaining high accuracy. This means that the model emulates medical practice in patients with the same characteristics as those described in the manuscript. The main limitation of our study is that data were generated from a population with acceptable glycemic control and thus, we do not know how these models will behave in other populations with different age ranges, higher HbA1c levels, or a high GV. Taking this into account, the RF technique could be a useful tool for clinical decision support systems on basal insulin dose for insulin pump users in our population. Another limitation was the dose prediction between 06:00–07:00 and 13:00–17:00. It could be explained by carbohydrate intake and bolus insulin. Also, physical activity and many regulatory hormonal secretions changes could affect glucose levels and basal insulin. It was our first approach to calculate basal insulin through artificial intelligence models. However, we need to work on the construction of models that predict basal and bolus insulin to improve prediction and get a better level of glycemic control. Anyway, the current error could be acceptable for basal insulin dose prediction but validation studies in other populations are necessary.

### Data Availability

Supporting information is available from the authors upon request by e-mail to guillermoeguzman@gmail.com.

### Conflicts of Interest

Oscar Bedoya performed as part of the employment at Universidad del Valle like as a professor. Other authors do not declare any potential conflicts of interest.

### References

- [1] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw, "IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030," *Diabetes Research and Clinical Practice*, vol. 94, no. 3, pp. 311–321, 2011.
- [2] International Diabetes Federation, *IDF Diabetes Atlas*, 9th edition, International Diabetes Federation, Brussels, Belgium, www.diabetesatlas.org., 9th edition.
- [3] I. N. Scobie, *Atlas of Diabetes Mellitus. 3 Edição*, Informa Healthcare, London, UK, 2007.
- [4] American Diabetes Association, "8. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes-2018," *Diabetes Care*, vol. 41, no. Suppl 1, pp. S73–S85, 2018.
- [5] F. Fatehi, L. C. Gray, and A. W. Russell, "Mobile health (mHealth) for diabetes care: opportunities and challenges," *Diabetes Technology & Therapeutics*, vol. 19, no. 1, pp. 1–3, 2017.
- [6] C. Hou, B. Carter, J. Hewitt, T. Francisa, and S. Mayor, "Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials," *Diabetes Care*, vol. 39, no. 11, pp. 2089–2095, 2016.
- [7] X. Liang, Q. Wang, X. Yang et al., "Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis," *Diabetic Medicine*, vol. 28, no. 4, pp. 455–463, 2011.
- [8] K. Pal, S. V. Eastwood, S. Michie et al., "Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus," *Cochrane Database of Systematic Reviews*, vol. 2013, no. 3, p. CD008776, 2013.
- [9] R. N. Angeles, M. I. Howard, and L. Dolovich, "The effectiveness of web-based tools for improving blood glucose control in patients with diabetes mellitus: a meta-analysis," *Canadian Journal of Diabetes*, vol. 35, no. 4, pp. 344–352, 2011.
- [10] F. Torrent-Fontbona, "Adaptive basal insulin recommender system based on kalman filter for type 1 diabetes," *Expert Systems with Applications*, vol. 101, pp. 1–7, 2018.
- [11] F. Torrent-Fontbona and B. López, "Personalized adaptive CBR bolus recommender system for type 1 diabetes," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 387–394, 2018.
- [12] G. Cappon, M. Vettoretti, F. Marturano, A. Facchinetti, and G. Sparacino, "A neural-network-based approach to personalize insulin bolus calculation using continuous glucose monitoring," *Journal of Diabetes Science and Technology*, vol. 12, no. 2, pp. 265–272, 2018.
- [13] R. Nimri, E. Dassau, T. Segall et al., "Adjusting insulin doses in patients with type 1 diabetes who use insulin pump and continuous glucose monitoring: variations among countries and physicians," *Diabetes, Obesity and Metabolism*, vol. 20, no. 10, pp. 2458–2466, 2018.
- [14] R. I. Khan, S. S. R. Nirzhor, A. M. Chowdhury, T. A. Shishir, and A. I. Khan, "A fuzzy logic based approach for the adjustment of insulin dosage for type 1 diabetes patients," *Therapy*, vol. 7, pp. 10–12, 2013.

- [15] C. C. Palerm, H. Zisser, L. Jovanovič, and F. J. Doyle, "A run-to-run control strategy to adjust basal insulin infusion rates in type 1 diabetes," *Journal of Process Control*, vol. 18, no. 3-4, pp. 258–265, 2008.
- [16] M. Ibbini, "A PI-fuzzy logic controller for the regulation of blood glucose level in diabetic patients," *Journal of Medical Engineering & Technology*, vol. 30, no. 2, pp. 83–92, 2006.
- [17] P. Herrero, P. Pesl, M. Reddy, N. Oliver, P. Georgiou, and C. Toumazou, "Advanced insulin bolus advisor based on run-to-run control and case-based reasoning," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 1087–1096, 2014.
- [18] B. López, C. Martin, and P. H. Viñas, "Special section on artificial intelligence for diabetes," *Artificial Intelligence in Medicine*, vol. 85, pp. 26-27, 2018.
- [19] A. T. M. Wasylewicz and A. M. J. W. Scheepers-Hoeks, "Clinical decision support systems," in *Fundamentals of Clinical Data Science*, pp. 153–169, Springer, Cham, Switzerland, 2019.
- [20] A. Ceriello, L. Monnier, and D. Owens, "Glycaemic variability in diabetes: clinical and therapeutic implications," *The Lancet Diabetes & Endocrinology*, vol. 7, no. 3, pp. 221–230, 2019.
- [21] Y.-H. Lin, Y.-Y. Huang, H.-Y. Chen et al., "Impact of carbohydrate on glucose variability in patients with type 1 diabetes assessed through professional continuous glucose monitoring: a retrospective study," *Diabetes Therapy*, vol. 10, no. 6, pp. 2289–2304, 2019.
- [22] J. Baron, H. McBain, and S. Newman, "The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review," *Journal of Diabetes Science and Technology*, vol. 6, no. 5, pp. 1185–1196, 2012.
- [23] B. Holtz and C. Lauckner, "Diabetes management via mobile phones: a systematic review," *Telemedicine and E-Health*, vol. 18, no. 3, pp. 175–184, 2012.
- [24] A. J. Deacon and S. Edirippulige, "Using mobile technology to motivate adolescents with type 1 diabetes mellitus: a systematic review of recent literature," *Journal of Telemedicine and Telecare*, vol. 21, no. 8, pp. 431–438, 2015.