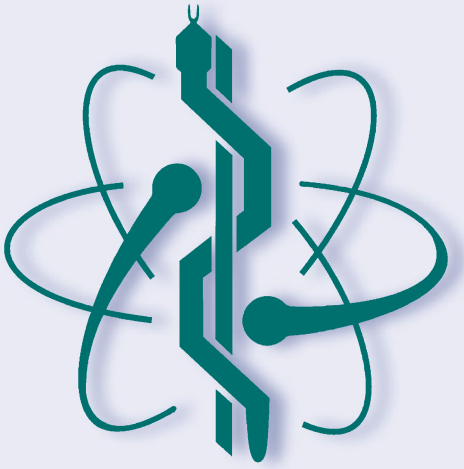
IFMBE Proceedings 108

Esteban Pino
Ratko Magjarević
Paulo de Carvalho *Editors*

International Conference on Biomedical and Health Informatics 2022

Proceedings of ICBHI 2022, November 24–26, 2022, Concepción, Chile



Springer

IFMBE Proceedings 108

The IFMBE Proceedings Book Series is an official publication of *the International Federation for Medical and Biological Engineering* (IFMBE). The series gathers the proceedings of various international conferences, which are either organized or endorsed by the Federation. Books published in this series report on cutting-edge findings and provide an informative survey on the most challenging topics and advances in the fields of medicine, biology, clinical engineering, and biophysics.

The series aims at disseminating high quality scientific information, encouraging both basic and applied research, and promoting world-wide collaboration between researchers and practitioners in the field of Medical and Biological Engineering.

Topics include, but are not limited to:

- Diagnostic Imaging, Image Processing, Biomedical Signal Processing
- Modeling and Simulation, Biomechanics
- Biomaterials, Cellular and Tissue Engineering
- Information and Communication in Medicine, Telemedicine and e-Health
- Instrumentation and Clinical Engineering
- Surgery, Minimal Invasive Interventions, Endoscopy and Image Guided Therapy
- Audiology, Ophthalmology, Emergency and Dental Medicine Applications
- Radiology, Radiation Oncology and Biological Effects of Radiation
- Drug Delivery and Pharmaceutical Engineering
- Neuroengineering, and Artificial Intelligence in Healthcare

IFMBE proceedings are indexed by SCOPUS, EI Compendex, Japanese Science and Technology Agency (JST), SCImago. They are also submitted for consideration by WoS.

Proposals can be submitted by contacting the Springer responsible editor shown on the series webpage (see “Contacts”), or by getting in touch with the series editor Ratko Magjarevic.

Esteban Pino · Ratko Magjarević ·
Paulo de Carvalho
Editors

International Conference on Biomedical and Health Informatics 2022

Proceedings of ICBHI 2022,
November 24–26, 2022, Concepción, Chile

Springer

Editors
Esteban Pino
Department of Electrical Engineering
Universidad de Concepcion
Concepcion, Chile

Paulo de Carvalho
Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal

Ratko Magjarević
Faculty of Electrical Engineering
and Computing
University of Zagreb
Zagreb, Croatia

ISSN 1680-0737 ISSN 1433-9277 (electronic)
IFMBE Proceedings
ISBN 978-3-031-59215-7 ISBN 978-3-031-59216-4 (eBook)
https://doi.org/10.1007/978-3-031-59216-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

This publication is the Proceedings of the International Conference on Biomedical and Health Informatics, ICBHI 2022, held in Concepción, Chile, from November 24–26, 2022, as the fifth in the series of topical conferences on biomedical and health informatics of the International Federation for Medical and Biological Engineering (IFMBE). The conference was jointly organized by the Digital Health Division (DHD) of the IFMBE, the Electrical Engineering Department at the University of Concepción and the Chilean Society of Health Informatics ACHISA, and it was endorsed by CORAL—the Latin American Regional Council on Biomedical Engineering. This conference is special for two reasons: it is the first IFMBE international conference in the field of BHI organized in Latin America and it is the first international conference organized by DHD after it achieved a significant and permanent position as a Division within IFMBE.

Implementation of digital technology in health care is often addressed as digital transformation aiming to increase the outcome and the quality of health care at each level. However, there are a lot of challenges in that process including capacity building, standardization, and interoperability within the health care system as well as security and privacy issues. Emerging technologies like big data and artificial intelligence are promising in terms of improving diagnostics and treatment, increasing efficiency, and reducing the costs of health care delivery. Therefore, it is important for scientists, researchers, and all those who work in development, application, management, and education and training in the health sector to hold gatherings like this conference, at the national level and for the whole of Latin America. At informal meetings of DHD members in conjunction with the conference, it was proposed to hold a regional Latin American BHI conference with the support of IFMBE every other year to encourage strengthening digital transformation in health care in the region.

For all those who are not familiar with the structure of the IFMBE, the change in status of a group within the Federation from a "working group" in health informatics and e-health to the Digital Health "Division" may not seem important. However, that change means that the activities and projects of the group have significantly increased from the time of its foundation in 2012 and that within the IFMBE-affiliated member societies, globally, there is a large interest for networking, exchange of ideas, research results, and cooperation in digital health. This change will not only increase the workload of the board members of the division, but it will also raise the expectations for the outcomes of the division's activity outcomes. Digital transformation is a global phenomenon, but it should be noticed that there are significant differences in the level of development and income, most often visible in individual regions of the world. The IFMBE DHD and CORAL will in future work systematically and build a wide network of scientists, professionals, and industry in the region of Latin America and the Caribbean in order to facilitate the implementation of digital technology and associated knowledge to health care in the region.

The described aspirations were also reflected in the program of the conference and in the content of its proceedings. The ICBHI 2022 Proceedings are divided into four parts:

Part I: Artificial intelligence and precision medicine
Part II: E-health and education
Part III: Medical devices and wearables technologies
Part IV: ICBHI Challenge: Ballistocardiogram beat detection.

While the first three parts bring some of the newest research and innovation mainly from the region, the fourth part presents papers of the finalists of the ICBHI 2022 Scientific Challenge. The Challenge is a competition meant for master and doctoral students, individuals or groups, who express their creativity in solving a clinically relevant open problem in digital health. The Challenge was established in 2017 by the HIeH working group and continued by the DHD. The best solutions in each competition are presented during the conference and the best solutions are awarded. The Challenge is supported by large databases that are offered to the DH community after the competition. In the fifth, ICBHI 2022 edition of the competition, the Challenge was in detection of heartbeats from Ballistocardiogram time series.

The conference in Concepción was held in the post-pandemic era but still suffered the consequences and travel restrictions. Therefore, it was organized in a hybrid mode, partially online. This proceedings volume covers selected presentations from the conference since not all authors were able to make the presentation. However, all submissions have been carefully and critically reviewed by at least two independent experts and additionally by at least one member of the scientific program committee. The editors are indebted to the acknowledged and highly experienced reviewers for their contribution to the quality of the conference publication. Both the ICBHI 2022 Conference and the publication of the proceedings by Springer Nature would not have been possible without the support and sponsorship of the IFMBE and its Digital Health Division. The editors are also grateful to the efforts of the local organizing committee members and their supporters for carefully and smoothly preparing and operating the conference. They especially thank all the team members from the University of Concepción for their dedication to the event.

Esteban Pino
Paulo de Carvalho
Ratko Magjarević

Organization

Organizers

Digital Health Division (DHD) of the IFMBE
Electrical Engineering Department at Universidad de Concepcion
Chilean Society of Health Informatics ACHISA

Conference Chair

Esteban J. Pino	Electrical Engineering Department, Universidad de Concepción, Chile

Conference Co-chairs

Paulo de Carvalho	University of Coimbra, Portugal
Ratko Magjarević	University of Zagreb, Croatia

Scientific Challenge Committee

Esteban Pino	Universidad de Concepción, Chile
Paulo de Carvalho	University of Coimbra, Portugal
Jorge Henriques	University of Coimbra, Portugal

Scientific Advisory Committee

Virginia Ballarin	Universidad Nacional de Mar del Plata, Argentina
Rosa Figueroa	Universidad de Concepción, Chile
Christopher Flores	Universidad de O'Higgins, Chile
Mauricio Hernández	Universidad de Antioquia, Colombia
Sandra Pérez	Pontificia Universidad Católica del Perú, Peru
Antonio Rienzo	Universidad de Valparaíso, Chile
Claudio Román	Universidad de Concepción, Chile
Marta Zequera	Pontificia Universidad Javeriana, Colombia

Local Organizing Committee

Fabian Alvarado	Universidad de Concepción, Chile
Francisca Zamorano	Universidad de Concepción, Chile
Macarena Araneda	Universidad de Concepción, Chile
María Fernanda Cornejo	Universidad de Concepción, Chile
Génesis González	Universidad de Concepción, Chile
Bastián Mercado	Universidad de Concepción, Chile
Carla Retamal	Universidad de Concepción, Chile
Benjamín Sepúlveda	Universidad de Concepción, Chile

Invited Speakers

Metin Akay	President of the IEEE Engineering Medicine and Biology Society, USA
Yasemin M. Akay	Department of Biomedical Engineering, Cullen College of Engineering, University of Houston, USA
Ricardo L. Armentano	Universidad de la República, Uruguay
Virginia Ballarin	Electronics Engineering Department of the Universidad Nacional de Mar del Plata, Argentina
Sonia Benitez	Jefe investigación e innovación tecnológica, Depto. Informática en Salud, Hospital Italiano, Argentina
Ratko Magjarević	President of the International Federation for Medical and Biological Engineering (IFMBE), Croatia
Nicos Maglaveras	Medical Informatics and Biomedical Imaging Technologies, Aristotle University of Thessaloniki, Thessaloniki, Greece
Carla Taramasco	Facultad de Ingeniería de la Universidad Andrés Bello, Chile
Carlos Zuñiga San Martín	Unidad Telenefrología, Hospital Digital, Ministerio de Salud, Chile

Sponsors

HealthTracker Analytics
CENIA—Centro Nacional de Inteligencia Artificial

Collaborators

CCB—Concepcion Convention Bureau
CENS—Centro Nacional en Sistemas de Información en Salud
CIV—Centro de Ingeniería para la Vida UdeC
CORAL—the Latin American Regional Council on Biomedical Engineering

Conference Webpage

www.icbhi2022.com

Contents

E-health and Education

Artificial Intelligence and Precision Medicine

Data-Driven Model for Long-Term Prediction of Blood Glucose in Type 2 Diabetes

Milene Jesus[1(✉)], Sara Zulj[2], Rogério T. Ribeiro[3], Marco Simões[1], Jorge Henriques[1], and Paulo Carvalho[1]

[1] Faculty of Science and Technology, University of Coimbra, Coimbra, Portugal
mijesus99@hotmail.com
[2] Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia
[3] Department of Medical Sciences, University of Aveiro, Aveiro, Portugal

Abstract. Type 2 diabetes mellitus (T2DM) is a disease that affects more than 380 million people worldwide. In this study, we developed a model, for these type of patients, that predicts blood glucose values over long prediction horizons (PHs), whose existence in the literature is almost nonexistent. These horizons allow patients to be warned in advance so that they can take action to avoid dangerous health situations. We used data from 3 of the 10 real patients available to test the implemented models. The overall results for the best model (simple Recurrent Neural Network) were: 34.82 mg/dL for root mean square error (RMSE) and 18.33% for mean absolute percentage error (MAPE) (PH = 2h); 46.59 mg/dL for RMSE and 24.35% for MAPE (PH = 4h).

Keywords: Type 2 diabetes mellitus · Glucose level prediction · CGM data

1 Introduction

Type 2 diabetes mellitus (T2DM), also called non-insulin-dependent diabetes, occurs due to a progressive loss of adequate insulin secretion from beta cells, often in the context of insulin resistance [1]. In 2017, the global prevalence of diabetes was 8.8% (in the 20–79 age group), representing 424.9 million individuals. By 2045, this figure is estimated to increase to 628.6 million people with diabetes [2]. About 90% of these cases correspond to patients with T2DM [2,3]. These values are mainly due to the increase in urbanization, population aging, obesity, unhealthy eating habits and sedentary lifestyle [1]. This pathology, when uncontrolled, has serious consequences, increasing the risk of cardiovascular and end-stage renal disease, retinopathy and neuropathy [1,3].

In an effort to keep glycemic values in range, these patients must follow a set of behavioral actions, such as sticking to a food plan, practicing sufficient physical activity and taking medication [3]. Some studies show that lifestyles

E. Pino et al. (Eds.): ICBHI 2022, IFMBE Proceedings 108, pp. 3–14, 2024.
https://doi.org/10.1007/978-3-031-59216-4_1

interventions are more effective than pharmacological ones and can even prevent cases of T2DM [3].

Blood glucose monitoring has been revolutionized in the last few decades by Continuous Glucose Monitoring devices (CGMs), which are temporary minimally invasive sensors inserted in subcutaneous tissue. These provide blood glucose (BG) readings every 1 to 5 min. Currently, there are intelligent computational techniques, such as Machine Learning (ML) and Artificial Intelligence (AI), that analyze and extract timely information for patients through the acquired data [4]. Some studies show that this combination of technologies (monitoring devices with intelligent algorithms) can contribute to the decrease in the value of Hemoglobin A1c (HbA1c) and improves glycemic control, self-efficacy, and self-care activities [5,6].

Several studies have been conducted on BG forecasting using only CGM data as input. Most models focus only on type 1 diabetes mellitus (T1DM), while literature regarding predictions in T2DM is scarce [7]. Martinsson et al. [8] used the Ohio T1DM Dataset for Blood Glucose Level Prediction to validate an long short-term memory (LSTM) model. They obtained RMSE values of 18.87 mg/dL for a 30 min prediction horizon (PH) - which specifies the target value that the algorithm should predict - and 31.40 mg/dL for a 60 min prediction horizon. A patient-specific prediction model based on LSTM was also trained and validated using the OhioT1DM dataset by Aliberti et al. [9]. The patient with the best predicted outcome out of the six patients had RMSE values of 11.55 mg/dL, 19.86 mg/dL, 25 mg/dL, and 30.95 mg/dL for 30, 45, 60, and 90 min. Zecchin et al. [10] proposed a predictor that combines a neural network model and a first-order polynomial extrapolation algorithm used in parallel to describe, respectively, the nonlinear and linear components of glucose dynamics. They monitored 15 Type-1 diabetic patients for 7 days using a CGM system that returns glucose values every minute. They showed that using carbohydrate intake information improves the accuracy of short-term prediction of glucose concentration. These deep learning based black-box models approaches carry significant limitations in interpretability, which becomes critical in algorithms that directly affect patient care. In order to overcome this difficulty, Zulj et al. [11] implemented case-based reasoning (CBR) for glucose prediction using CGM data. The study was conducted using data from 20 subjects recorded under free-living conditions. The best models developed by the authors achieved a mean absolute error (MAE) of 13.35 mg/dL for PH = 30 min and 30.23 mg/dL for PH = 60 min.

Regarding the prediction of glucose values for T2DM patients, one study was found for hospitalized patients with this disease. Kim et al. [12] collected data from 20 patients for one week on a CGM device. The model used the last 35 min to predict blood sugar for the next 30 min. The best model developed, using gated recurrent unit (GRU), obtained an RMSE of 21.5 mg/dL and a mean absolute percentage error (MAPE) of 11.1%. Other studies aiming to predict T2DM outbreak have been found [13,14]. Although they are performed in patients with T2DM they differ from our study, which predicts blood glucose with CGM data.

To our knowledge, there are no studies on the prediction of blood glucose over very long prediction horizons, these rarely exceed 1h, in either type 1 or type 2 diabetes patients.

The main goal of our study is to develop a model that predicts clinically relevant long-term blood sugar values for patients with type 2 diabetes. This prediction is based solely on the patients' CGM history. These patients may benefit from longer predictions, such as for 2h, 4h or 12h prediction horizons, since these have the ability to reflect their everyday behaviors, such as eating a meal or exercising.

2 Materials and Methods

2.1 Dataset

In the present work a dataset provided by Associação Protetora dos Diabéticos de Portugal (APDP), with real patient data, was used to validate the models implemented. Blood sugar levels were recorded in free-living conditions using the Medtronic iPro2 CGMs with a 5-minute sampling period. The CGM time-series containing the blood glucose concentration levels were collected as part of an observational research that included adult participants with type 1 and type 2 diabetes mellitus who were receiving hemodialysis. The duration of the CGM time-series ranged from 2 days to 8 days, had cutoffs for values below 40 mg/dL and above 400 mg/dL, which corresponds to the sensor range, and may include several periods of missing data.

For the experiments, anonymized data from 10 subjects were selected from the larger dataset based on the following criteria: 1) Have type 2 diabetes mellitus, 2) Participate in two experiments. Following the selection criteria, we derived the new dataset that includes 10 T2DM subjects, each represented by variable sizes of the CGM time series. Descriptive statistics for the selected subjects are presented in Table 1, and an overview of the CGM profiles is presented in Table 2.

Table 1. Descriptive statistics of the dataset.

Characteristic	Mean ± SD
Gender	5F/5M
Age (years)	73.7 ± 7.4
Diabetes duration (years)	16.7 ± 8.2
Body mass index (BMI) (kg/m2)	31.8 ± 2.8
Fat mass (%)	44.0 ± 5.1
HbA1c (%)	7.4 ± 1.8

Table 2. Glucose profile statistics of the dataset per trial.

	Trial No.1	Trial No.2	Total
Average glucose concentration (mg/dL)	175.8 ± 58.3	163.8 ± 40.7	169.8 ± 49.5
Minimal glucose concentration (mg/dL)	63.9 ± 19.1	52.4 ± 17.2	58.1 ± 18.1
Maximal glucose concentration (mg/dL)	310.7 ± 99.1	322.4 ± 70.1	316.5 ± 84.6
% hypoglycemic values ($\leq$70 mg/dL)	2.7 ± 5.2	3.1 ± 2.4	2.9 ± 3.8
% values in healthy range	55.9 ± 32.7	62.3 ± 24.6	59.1 ± 28.7
% hyperglycemic values ($\geq$180 mg/dL)	41.4 ± 34.5	34.5 ± 25.6	38.0 ± 30.0

2.2 Methods

As preprocessing, missing values were treated and the data were normalized. For the latter we used data scaling based on min-max normalization between values $[-1, 1]$.

We transformed the CGM records into a supervised learning problem by framing the data using the sliding window method (with a step size of 1 generating each sample). Each example in supervised learning is a pair that includes an input item and the desired output value. The number of prior values that must be used simultaneously by the algorithm is called the lookback (L). A grid search was performed to obtain which lookback is optimal per PH, that is, the one that, using the selected evaluation criteria, produces values that are the most satisfactory. For PHs equal to 2h, 4h and 12h the lookbacks that obtained the best results were 12h, 12h and 24h, respectively.

When performing the hyperparameter tuning tests, some difficulties arose in the prediction task. The continuous use of all measurements to determine only one exact value after a few hours becomes very demanding for the models. The dynamics of glucose is complex and depends on various factors such as diet, exercise, hormonal values, psychological stress, etc. For longer horizons where the intervention of the above mentioned factors is accentuated, it becomes difficult to achieve satisfactory results. Therefore, since the aim of this study is to predict BG values for T2DM patients and not for T1DM patients (who due to their dependence on insulin and medication need continuous and rigorous predictions), a different approach was created to input the data into the models. In this, the points of the CGM record were grouped by the average every 24 points, corresponding to every 2h. This aims to prioritize glucose trends rather

than its exact value. Here, the crucial point is not to tell the patient the exact glucose value within 4h, but to be able to predict whether the blood glucose value will tend to rise or fall to dangerous values, so that the patient on being alerted can have some kind of reaction that prevents this outcome.

So, Table 3 illustrates an example of how the data is framed for the forecast. For a forecasting horizon of 4h, i.e., to predict the average of two 2h blocks after the last input value, the averages of 6 previous 2h blocks (12h) are needed as inputs.

Table 3. Form of the input sample. Example where x is a CGM time series with the values grouped by the average every 2h, where the lookback corresponds to 12h and PH = 4h.

Input values					PH	Target value
x_1	x_2	x_3	...	x_6	...	x_8
x_2	x_3	x_4	...	x_7	...	x_9
x_3	x_4	x_5	...	x_8	...	x_{10}
...						

The prediction module was built with the high-level neural networks API Keras version 2.7.0 in the Python 3.9.7 environment. Six prediction models were implemented to determine the best one for the task of forecasting blood sugar level at the 2h, 4h, and 12h PHs. The selected algorithms were Autoregressive Integrated Moving Average (ARIMA), CBR and four neural networks: simple recurrent neural network (RNN), GRU, LSTM and Jump Neural Network (JNN), since they stood out in the literature review for better performance compared to other models.

The ARIMA model was implemented as a baseline, since this model is considered to be one of the most flexible and popular autoregressive techniques for continuous time series forecasting [15].

CBR is a methodology based on the intuition that similar problems often have similar solutions. It provides an inherent model-specific approach to interpretability [11]. A *case* is represented as the ordered pair (*problem, solution*) [11]. As mentioned earlier, in this project the CGM data is organized as pairs. Thus, a *case base* is created with all *cases* from all training patients. The solution is learned from the set of existing instances in the *case base* for each new instance of the problem [16]. The proposed CBR model described in [11] was implemented. All the steps described in the proposed method were followed, so we suggest the reader to analyse the point regarding Methods from the original article for a better understanding.

Relatively to the neural network models, the JNN proposed in [10] was implemented. Four neurons were used in the hidden layer and as input only the blood glucose concentration values were considered, disregarding the input absorption model. The implementation with this limitation on input was performed by the

same author in [17]. Thus, for a detailed description of the network architecture, we would like to refer the reader to the appendix found in the original paper [17]. For the remaining neural networks, all implementations consisted of a single hidden layer. In each network 50 units were used in the hidden layers. A dense layer of one unit was also used to produce the final predicted blood glucose value. The rest of the network parameters (number of epochs, batch size, optimization function, loss function and learning rate) were selected from a grid search. The results are shown in Table 4. This table presents the most common parameter choices across the different training sets used.

Table 4. Grid search results for the RNN, GRU and LSTM hyperparameters.

Hyperparameter	Most common option
Number of epochs	300
Batch size	8
Optimization function	ReLu
Loss function	MSE
Learning rate	3e − 5

2.3 Evaluation Metrics

Empirical accuracy of the model was evaluated using RMSE and MAPE between predicted time-series $\hat{y}$ by and target time-series y:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{1}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|\,\hat{y}_i - y_i\,|}{y_i} * 100\% \tag{2}$$

To measure the clinical accuracy of the models' predictions the Clarke Error Grid (CEGA) was used. Errors are divided into zones in this grid system, and each zone is made up of a range of reference and forecast values. On the error grid, the correspondence between real and predicted blood glucose levels is displayed. Each of these pairs falls into one of the error grid zones. Zones A and B are completely appropriate in terms of the therapeutic setting.

3 Results

3.1 Analytical Evaluation

The subjects in the APDP dataset were randomly divided into 70% of the data for training and 30% for testing. That is, 7 subjects (14 trials) constituted the

training set and the remaining 3 subjects (6 trials) constituted in the test set. We show the results of both trials (marked with an identification code 102, 106, and 116).

Tables 5 and 6 report the experimental results obtained by running the final models for PH = 2h, using L = 12h, evaluated by RMSE and MAPE, respectively.

Table 5. Comparison of the performance for PH = 2h - RMSE.

RMSE (mg/dL) - PH = 2h							
Subject ID	Trial No.	ARIMA	CBR	RNN	GRU	LSTM	JNN
102	1	26.81	22.89	22.74	23.75	23.71	24.67
102	2	45.28	40.19	41.64	41.02	41.09	43.36
106	1	50.11	36.85	34.43	35.42	35.68	37.04
106	2	44.63	31.87	30.68	30.71	30.84	33.20
116	1	46.84	33.46	32.76	32.54	32.58	32.37
116	2	66.01	48.75	46.68	47.74	47.87	51.91
	Mean	46.61	35.67	34.82	35.20	35.29	37.09
	SD	11.45	7.91	7.69	7.63	7.67	8.67

Table 6. Comparison of the performance for PH = 2h - MAPE.

MAPE (%) - PH = 2h							
Subject ID	Trial No.	ARIMA	CBR	RNN	GRU	LSTM	JNN
102	1	18.01	17.30	16.03	16.38	16.41	16.97
102	2	28.08	25.88	27.49	26.50	26.31	27.01
106	1	24.32	17.71	17.03	17.42	17.56	18.20
106	2	20.31	16.27	16.02	15.99	15.20	16.66
116	1	19.60	12.98	13.47	13.06	12.87	12.96
116	2	29.14	19.40	19.92	19.77	19.90	21.38
	Mean	23.24	18.26	18.33	18.19	18.04	18.86
	SD	4.26	3.92	4.51	4.21	4.27	4.41

Tables 7 and 8 report the experimental results obtained by running the final models for PH = 4h, using L = 12h, evaluated by RMSE and MAPE, respectively.

Analyzing the results, it can be seen that the CBR and JNN shown overall better results than the ARIMA model, but not as satisfactory as the other three neural networks. The RNN, GRU and LSTM models presented very close results at both horizons. As for PH = 12h, it showed lower results than the others. This is due to the fact that glucose dynamics depends on several parameters, and in a 12h horizon there may be several episodes that influence the patient's glucose values.

Table 7. Comparison of the performance for PH = 4h - RMSE.

RMSE (mg/dL) - PH = 4h							
Subject ID	Trial No.	ARIMA	CBR	RNN	GRU	LSTM	JNN
102	1	26.75	27.89	27.15	25.84	25.36	27.63
102	2	49.99	50.40	49.20	48.85	48.04	48.49
106	1	56.87	48.91	47.13	47.55	47.16	49.35
106	2	54.79	44.11	41.75	42.02	44.78	47.64
116	1	56.08	47.27	47.73	46.90	48.46	47.16
116	2	76.15	69.97	66.59	67.73	69.28	70.43
	Mean	53.44	47.97	46.59	46.48	47.18	48.45
	SD	14.50	12.09	11.62	12.27	12.73	12.38

Table 8. Comparison of the performance for PH = 4h - MAPE.

MAPE (%) - PH = 4h							
Subject ID	Trial No.	ARIMA	CBR	RNN	GRU	LSTM	JNN
102	1	17.87	20.88	19.12	18.62	17.91	19.36
102	2	31.71	32.27	31.66	31.06	30.61	29.03
106	1	29.51	24.21	24.09	24.46	24.33	25.33
106	2	27.78	23.06	22.35	22.14	23.49	23.75
116	1	24.05	17.96	19.51	19.40	19.59	19.06
116	2	34.80	30.14	29.37	31.19	32.23	31.70
	Mean	27.62	24.76	24.35	24.48	24.69	24.70
	SD	5.47	5.00	4.71	5.06	5.25	4.64

The results of this study are presented using two trials for each of the three test subjects. This allows us to gain insight into inter- and intra-subject variations. By analyzing the MAPE values we can get a better idea of these variations. Patient 102 shows the largest differences between trials for the 2h and 4h prediction horizons. There is an increase in the error by ∼10% and ∼23% from the first to the second trial, respectively.

3.2 Clinical Evaluation

Despite the fact that the metrics mentioned above are crucial for comprehending the performance and prediction accuracy of different models from a regression analysis point of view, they are unable to identify the most significant outliers and do not offer any details about the clinical impact of prediction errors and their effects on medical treatment decisions. Therefore, we combined our assessment with CEGA analysis to present a more full view of the models' performance.

The comparison results for PH = 2h and 4h for a percentage of predictions falling into zones A and B of the error grid analysis are in Tables 9 and 10,

Table 9. Comparison of the performance for PH = 2h - Grid error analysis, zones A and B.

Zone A, zone B (%) - PH = 2h							
Subject ID	Trial No.	ARIMA	CBR	RNN	GRU	LSTM	JNN
102	1	67.90, 30.86	61.73, 37.04	69.14, 29.63	66.67, 32.10	66.67, 32.10	64.20, 34.57
102	2	42.50, 55.00	46.25, 52.50	42.50, 56.25	42.50, 56.25	38.75, 60.00	45.00, 53.75
106	1	52.5, 45.00	71.25, 28.75	68.75, 31.25	66.25, 33.75	68.75, 31.25	67.50, 32.50
106	2	53.16, 43.04	64,56, 34,18	68.35, 30.38	68.35, 30.38	67.09, 31.65	68.35, 30.38
116	1	60.34, 37.93	84,48, 15,52	82.76, 17.24	84.48, 15.52	86.21, 13.79	82.76, 32.91
116	2	44.30, 44.30	68,35, 26,58	68.35, 26.58	68.35, 26.58	68.35, 26.58	62.02, 32.91
	Mean	53.45, 42.69	66,10, 32,43	66.64, 31.89	66.10, 32.43	65.97, 32.56	64.97, 32.91
	SD	8.76, 7.33	11,44, 11,26	11.97, 11.86	12.28, 12.20	13.94, 13.81	11.12, 10.70

Table 10. Comparison of the performance for PH = 4h - Grid error analysis, zones A and B.

Zone A, zone B (%) - PH = 4h							
Subject ID	Trial No.	ARIMA	CBR	RNN	GRU	LSTM	JNN
102	1	67.50, 32.50	57.50, 42.50	62.50, 37.50	60.00, 40.00	65.00, 35.00	65.00, 35.00
102	2	43.04, 54.43	39.24, 58.23	37.97, 59.49	45.57, 51.90	41.77, 55.70	45.57, 51.90
106	1	46.83, 48.10	54.43, 41.77	58.23, 40.51	54.43, 44.30	51.90, 46.83	51.28, 42.31
106	2	48.72, 43.59	57.69, 38.46	57.69, 39.74	57.69, 38.46	61.54, 32.05	61.40, 36.84
116	1	49.12, 47.37	63.16, 33.33	57.89, 42.10	56.14, 38.46	59.65, 40.35	42.31, 50.00
116	2	34.61, 48.72	47.44, 43.59	42.31, 50.00	42.31, 50.00	41.03, 51.28	52.07, 44.45
	Mean	48.30, 45.78	53.24, 42.98	52.77, 44.89	52.69, 43.86	53.48, 43.53	52.07, 44.45
	SD	9.89, 6.74	7.82, 7.62	9.16, 7.61	6.48, 4.86	9.40, 8.50	8.37, 6.78

respectively. For each patient in both trials, the percentages of predictions falling into zones A and B, separated by a comma, are shown.

All models tested performed satisfactorily. More than 96% of the data in the 2h prediction horizon fall into zones A and B. For the 4h prediction horizon, the percentage of values in the clinically acceptable zones decreases to 95%. Consistently with analytical assessment, the performance got worse when increasing further the prediction horizon. The ARIMA model presented, again, the worst

results. For a prediction interval of 2h and 4h, the CBR model presented some of the best results together with the neural networks.

3.3 Discussion

By taking into consideration the balance between performance and time cost for pre-train, the RNN model is chosen as the final model for this module. An associated complexity, and consequent computational burden, of the GRU and LSTM models do not contribute significantly to the improvement of the results.

There are large differences in prediction between subjects (inter-subject variability), which suggests that personalized models, i.e., using the patient's own historical glucose data, would be able to achieve better results. Furthermore, since there was also large variability between trials for each patient (intra-subject variability), it could suggest the need for the algorithms to relearn the parameters as time progresses to overcome dynamic changes in the subject's glucose. Using only past and present glucose values as input data do not portray the complexity of BG dynamics. Adding data from other sources and viable sensors that measure variables affecting the metabolic process could lead to optimized results. These information could be about food intake, insulin injections, exercise, and mental health-related parameters such as stress levels.

These experiments have a number of notable drawbacks. First, we are conscious of the big bias presented by the fact that the dataset size of 10 participants is regarded as small. Second, the dataset was not evaluated in any way as a representation of the dynamics of the general population. The process for expanding the dataset with new participants should be investigated in greater detail in order to enhance the model.

4 Conclusion

In this study we addressed the problem of glucose level forecasting, using only CGM data as input, for T2DM patients. Our specific goal was to use a multi-patient training set to create a generalizable model for glucose level prediction that may be used to forecast future glucose levels for a new patient. This makes it possible to increase the models' usefulness even when they are just based on prior patient records.

It is concluded that the implementation of the RNN model can be used achieving satisfactory results for the 2h and 4h forecast horizon. Its use in a 12h forecast horizon does not show satisfying results. The global results of the 3 patients for the final model were: 34.82 mg/dL for RMSE and 18.33% for MAPE (PH = 2h) and 46.59 mg/dL for RMSE and 24.35% for MAPE (PH = 4h).

A contribution of this work is that we have developed a prediction model for patients with type 2 diabetes, which is scarce in the literature. Further studies in this area are needed to identify which prediction horizons are most useful and to improve models for patients with T2DM. We believe that these predictions can be a good basis to support a recommendation system that infers about diet and exercise in patients with T2DM.

Acknowledgements. This work is funded by the project POWER (grant number POCI-01-0247-FEDER-070365), co-financed by the European Regional Development Fund (FEDER), through Portugal 2020 (PT2020), and by the Competitiveness and Internationalization Operational Programme (COMPETE 2020). We would like to thank Rita Andrade from APDP - Diabetes Portugal, Education and Research Centre (APDP-ERC) for providing the database used in this project. The development of this study took place at the Center of Informatics and Systems of the University of Coimbra (CISUC).

Conflict of Interest. All authors of this article declare no conflict of interest. The decision to publish the results of the study was made by the authors without the funders' interference.

References

1. Association American Diabetes, et al.: 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2021. Diabetes Care **44**, S15–S33 (2021)
2. Standl, E., Khunti, K., Hansen, T.B., Schnell, O.: The global epidemics of diabetes in the 21st century: current situation and perspectives. Eur. J. Prev. Cardiol. **26**, 7–14 (2019)
3. Zheng, Y., Ley, S.H., Hu, F.B.: Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat. Rev. Endocrinol. **14**, 88–98 (2018)
4. Chaki, J., Ganesh, S.T., Cidham, S.K., Theertan, S.A.: Machine learning and artificial intelligence based diabetes mellitus detection and self-management: a systematic review. J. King Saud Univ. Comput. Inf. Sci. (2020)
5. Ellahham, S.: Artificial intelligence: the future for diabetes care. Am. J. Med. **133**, 895–900 (2020)
6. Wu, I.X.Y., Kee, J.C.Y., Threapleton, D.E., et al.: Effectiveness of smartphone technologies on glycaemic control in patients with type 2 diabetes: systematic review with meta-analysis of 17 trials. Obes. Rev. **19**, 825–838 (2018)
7. D'Antoni, F., Merone, M., Piemonte, V., Iannello, G., Soda, P.: Auto-regressive time delayed jump neural network for blood glucose levels forecasting. Knowl.-Based Syst. **203**, 106134 (2020)
8. Martinsson, J., Schliep, A., Eliasson, B., Meijner, C., Persson, S., Mogren, O.: Automatic blood glucose prediction with confidence using recurrent neural networks. In: Khd@ ijcai (2018)
9. Aliberti, A., Bagatin, A., Acquaviva, A., Macii, E., Patti, E.: Data driven patient-specialized neural networks for blood glucose prediction. In: 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1–6. IEEE (2020)
10. Zecchin, C., Facchinetti, A., Sparacino, G., De Nicolao, G., Cobelli, C.: Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. IEEE Trans. Biomed. Eng. **59**, 1550–1560 (2012)
11. Zulj, S., Carvalho, P., Ribeiro, R.T., Andrade, R., Magjarevic, R.: Data size considerations and hyperparameter choices in case-based reasoning approach to glucose prediction. Biocybern. Biomed. Eng. **41**, 733–745 (2021)
12. Kim, D.-Y., Choi, D.-S., Kim, J., et al.: Developing an individual glucose prediction model using recurrent neural network. Sensors **20**, 6460 (2020)

13. Wu, H., Yang, S., Huang, Z., He, J., Wang, X.: Type 2 diabetes mellitus prediction model based on data mining. Inform. Med. Unlocked **10**, 100–107 (2018)
14. Kazerouni, F., Bayani, A., Asadi, F., Saeidi, L., Parvizi, N., Mansoori, Z.: Type2 diabetes mellitus prediction using data mining algorithms based on the long-noncoding RNAs expression: a comparison of four data mining approaches. BMC Bioinform. **21**, 1–13 (2020)
15. Mohebbi, A., Johansen, A.R., Hansen, N., et al.: Short term blood glucose prediction based on continuous glucose monitoring data. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 5140–5145. IEEE (2020)
16. Choudhury, N., Begum, S.A.: A survey on case-based reasoning in medicine. Int. J. Adv. Comput. Sci. Appl. **7**, 136–144 (2016)
17. Zecchin, C., Facchinetti, A., Sparacino, G., Cobelli, C.: How much is short-term glucose prediction in type 1 diabetes improved by adding insulin delivery and meal content information to CGM data? A proof-of-concept study. J. Diab. Sci. Technol. **10**, 1149–1160 (2016)

Semantic of Automatically Generated Interval-Valued Memberships Functions in Brain Magnetic Resonance Images

D. S. Comas[1,2(✉)], G. J. Meschino[3], and V. L. Ballarin[2]

[1] Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Mar del Plata, Argentina
diego.comas@fi.mdp.edu.ar

[2] Image Processing Lab, Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (ICyTE), Universidad Nacional de Mar del Plata-CONICET, Mar del Plata, Argentina

[3] Bioengineering Lab, Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (ICyTE), Universidad Nacional de Mar del Plata-CONICET, Mar del Plata, Argentina

Abstract. Medical image segmentation plays a crucial role in diagnosis assistance. In previous works, we proposed a classification method called Type-2 Label-based Fuzzy Predicate Classification (T2-LFPC), which generates Interval-Valued Membership Functions (IVMF) and fuzzy predicates. They can be analyzed to interpret the images. In this work, a methodology is proposed to study the semantic of IVMF generated from brain MRI as input of the T2-LFPC. It is possible to understand both membership functions and predicates by visual inspecting positions and shapes of the IVMF. Some changes are applied on the images. Transformations include: zero mean additive noise, contrast-stretching and brightness increase and decrease. Changes in the images by transformations are reflected in the histograms of the pixels belonging to white matter, gray matter, and cerebrospinal fluid, in the IVMF and the values of their measures. Therefore, as changes are reflected in the IVMF as expected, the methodology proposed here could be considered suitable for image analysis.

Keywords: Brain magnetic resonance image segmentation · Semantic · Knowledge discovery · Interval-valued fuzzy membership functions

1 Introduction

Medical images play a crucial role in diagnosis assistance [1]. The technological advance of the past decades has substantially increased the available information, so emerging new processing techniques [1, 2]. As a consequence, new segmentation problems are constantly generated [3]. Therefore, any method which not only solves the segmentation but, also, allows to discover interpretable knowledge, expressed in a human language, can lead to significant contributions to the study and solution of certain medical problems [1].

In previous works [4, 5], we proposed a classification method called Type-2 Label-based Fuzzy Predicate Classification (T2-LFPC), which can be used for image segmentation when a Gold-Standard is provided. It consists of four stages: A) random partition

E. Pino et al. (Eds.): ICBHI 2022, IFMBE Proceedings 108, pp. 15–23, 2024.
https://doi.org/10.1007/978-3-031-59216-4_2

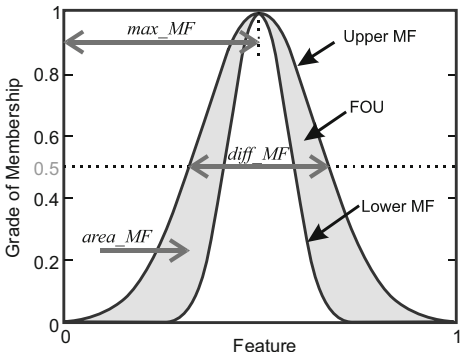


Fig. 1. Definition of the three measures proposed for analyzing interval-valued membership functions.

of the data, B) extraction of class prototypes, C) generation of a fuzzy predicates system, and D) optimization. The stages A and B are focused in providing prototypes for each label in the Gold-Standard capturing collections of common attributes in the data in each label. The stage C is focused on generating both membership functions and predicates, relating attributes of the features with properties observed in the prototypes. Besides classification, the method automatically generates Interval-Valued Membership Functions (IVMF) and fuzzy predicates, which can be analyzed to interpret the images. In these works, we also proposed some visual methods to perform the analysis. In the present paper, we focus on brain PD, T1, and T2 weighted Magnetic Resonance Images (MRI) to continue the analysis, considering new relevant aspects and tests. The main contributions are: a) analysis of the effects of zero mean additive noise, global contrast and brightness transformations, and noise in the acquisition channel (in simulated images); b) analysis of the effects of these transformations on the tissues histograms for each sequence; c) link the histograms changes with changes in the IVMF and in the segmentation performance; d) definition of a general criteria connecting the transformations applied to the input images with changes in the functions generated by the T2-LFPC.

2 Proposed Approach

An IVMF is bounded by two type-1 membership functions called Lower Membership Function (LMF) and Upper Membership Function (UMF) [6]. The area between the LMF and the UMF is called Footprint of Uncertainly (FOU) which is related to the vagueness or imprecision around the attribute described [6–8].

As mentioned, it is possible to understand both membership functions and predicates generated by the T2-LFPC method by visual inspecting position and shapes of the IVMF. In the present work the next three measures on interval membership functions are introduced (see Fig. 1):

i. *max_MF*: Position of the maximum of the IVMF. It is the value of the feature who better satisfy the attribute of the label.

ii. *diff_MF*: Width of the IVMF where it is equal to 0.5, associated to the imprecision for describing the values related to a label: small, medium, and high width means low, medium, and high imprecision, respectively.
iii. *area_MF*: Area of the IVMF: small area = low vagueness (it is possible to be more precise when explaining the grade in what the attribute is met by the feature), medium area = average vagueness, and large area = high vagueness.

The previously defined measure provides a way of quantify and describe attributes associated to IVMF as well as imprecision and vagueness. If IVMF are automatically generated from data (as it is the case of the T2-LFPC method), the conclusions of analyzing the values of the measures can be extended to the data used as input of the T2-LFPC.

Two brain MRI datasets were using, with pixel classified in three classes: White Matter (WM), Gray Matter (GM) and Cerebrospinal Fluid (CSF). The datasets used are:

- *Dataset #1*: 10000 pixels per tissue randomly selected from real brain MRI (30000 data, 3 classes, 3 features), acquired at the Dementia Clinic of the Institute for Neurological Research "Raúl Carrea" (Buenos Aires, Argentina) with a 1.5 T system with protocol: coronal 3D T1-weighted gradient echoes orthogonal to the AC-PC line (TR/TE = 24/5 ms, slice thickness = 1.5 mm); coronal proton density (PD); T2-weighted fast spin echoes oriented (TR/TE1/TE2 = 3,500/32/96 ms, echo train length = 8, slice thickness = 3 mm). In order to have a Gold Standard, pixels were classified using BRAINS [9] and optimized by medical experts.
- *Dataset #2:* Simulated brain MRI [10], 4000 pixels randomly selected per class (12000 data, 3 classes, 3 features). Data were taken without any distortion and were generated by computer simulation.

The next methodology is proposed to study the semantic of IVMF generated from brain MRI as input of the T2-LFPC:

a) Define a dataset of PD, T1, and T2 weighted brain MRI and their Gold Standard.
b) Apply changes on the images (PD, T1, and T2). In the case of dataset #1, transformations include: zero mean additive noise, contrast-stretching (considering a linear transformation mapping the minimum level of gray to 0 and the maximum to 255) and brightness changes: brightness increase (adding an offset mapping the maximum to 255 and so on), and brightness decrease (subtracting an offset mapping the minimum to 0 and so on). As noted, the brightness changes do not affect the contrast. The dataset #2 was selected as it provides a simulation of noise in the acquisition channel.
c) Apply the T2-LFPC method to each dataset, computing segmentation performance.
d) Study visually the IVMF generated for each variant and compute the measures previously introduced in order to analyze: relative position in the scale of the feature (discovering the attribute), area (describing vagueness around the attributes), and width (describing the spread of the data in the label).
e) Compute the histograms of the pixel belonging to each tissue in each variant to link the transformations applied with changes in the histograms and in the IVMF.

The methods T1-LFPC (a variant of the T2-LFPC using type-1 membership functions), Probabilistic Neural Networks (PNN) [11], Multi-Layer Perceptrons (MLP) [12],