DATA DRIVEN PATIENT-SPECIALIZED NEURAL NETWORKS FOR BLOOD GLUCOSE PREDICTION

Alessandro Aliberti[1]*, Andrea Bagatin*[1]*, Andrea Acquaviva*[2]*, Enrico Macii*[1] *and Edoardo Patti*[1]

[1]Politecnico di Torino, Torino, Italy - [2]University of Bologna, Bologna, Italy

ABSTRACT

Diabetes is an autoimmune disease characterized by glucose levels dysfunctions. It involves continuous monitoring combined with insulin treatment. Nowadays, continuous glucose monitoring systems (CGMs) have led to a greater availability of data. These can be effectively used by machine learning techniques to infer future values of the glycaemic concentration, allowing the early prevention of dangerous states and a better optimisation of the diabetic treatment. In this work, we investigate a patient-specialized prediction model. Thus, we designed a specialized solution based on Long Short-Term Memory (LSTM) neural network. Our solution was experimentally compared with two literature approaches, respectively based on Feed-Forward (FNN) and Recurrent (RNN) neural networks. The experimental results have highlighted that our LSTM solution obtained good performance both for short- and long-term glucose level inference (60 min.), overcoming the other methods both in terms of correlation between measured and predicted glucose signal and in terms of clinical outcome.

Index Terms— CGM, Diabetes, Neural Networks, Time-series Analysis, Machine Learning, LSTM.

1. INTRODUCTION

Diabetes is a chronic disease characterized by imbalances in blood glucose levels due to dysfunctional insulin production. This pathology is caused by antibodies that attack the β cells, which are responsible for the production of insulin in the pancreas. Defects in either insulin secretion or insulin sensitivity (or both) can lead to an increase in blood glucose beyond physiological levels. There are three main categories of diabetes. Type I diabetes, resulting in a loss of insulin, affects about 10% of the diabetic patients since childhood. Type II diabetes, related to an insulin insensitivity or inhibition, represents most common form of diabetes and usually develops in adulthood. Then, the gestational diabetes that refers to a temporary condition occurring in woman during pregnancy. In this comparative analysis, we focus on Type I diabetes. This disease, widespread worldwide, involves recurrent monitoring of blood glucose levels together with insulin treatment. There are mainly two types of monitoring systems i) the self-monitoring, performed by combining a finger and a measuring device, and ii) the continuous glucose monitoring system (CGM). Thanks to the increasing availability of low-cost devices the medical sector is moving towards the concept of *smart healthcare* [1] leading to greater availability of data and thus stimulating research to develop techniques for predicting future values within different prediction time horizons. This enables new and robust methodologies chasing two-fold benefits: i) preventing potentially dangerous complications (i.e. hyper- or hypoglycaemic states) and ii) optimizing the insulin dose that needs to be injected. Therefore, the patient can be subjected to continuous remote monitoring by the primary care physicians, triggering automatic alert mechanisms and, whenever needed, faster hospitalisation procedures [2].

Nowadays, sensor-based user behaviour and health status monitoring are getting much attention [3]. In detail, since the introduction of CGM devices, literature has proposed several approaches for glucose level prediction. Generally, these are divided into the two macro-categories: i) approaches based physiological models, reproducing metabolic processes of a patient by means of equations that mathematically describe glucose kinetics [4] and ii) data-driven approaches, that infer the future values of glucose concentration by applying machine learning techniques [5]. However, machine learning techniques are generally preferred because they promise higher flexibility and capability w.r.t. fixed physiological model. Indeed, they are immune to unpredictable variability of glucose kinetics due to either internal (e.g. different device calibrations) or external factors (e.g. physical activities, sudden stress, etc.). Also they do not depend on fixed parameters. In recent years, the most common approaches use autoregressive models (AR) or artificial neural networks (ANN) [6]. However, this method suffers from significant prediction errors and a very limited forecasting window (i.e. about 30 min.). Better accuracy values based on ANN or on SVR techniques are shown by the most recent works [7, 8]. Further developments are presented by [9], using a model based on feed-forward ANN, and by [10], using a recurrent neural network (RNN). However the prediction time horizons are still modest. While the most consolidated works are generally based on shallow neural networks, few recent studies started proposing deep learning techniques [11, 12]. Nevertheless, these methods i) are very demanding, ii) need a huge

set of data for an effective training iii) are not easily interpreted.

In this work, we propose a patient-specialized blood glucose prediction system exploiting a data-driven approach. Thus, we designed and implemented our solution based on Long Short-Term Memory (LSTM) neural network that is generally acknowledged as one of the best architecture for time-series predictions thanks to its versatility and flexibility [13]. Consequently, we tested and evaluated our solution capabilities and performances to improve the prediction accuracy, possibly on a much larger forecasting time horizon. For a fair validation of our methodology, we exploited the very same dataset for training and testing the two most significant state-of-art neural networks in blood glucose prediction (i.e. [9] and [10]). Then, to evaluate the performance of the models, we exploited different quantitative and qualitative performance indexes to identify the most promising. The solution proposed in this work is part of a broader framework introduced in our previous work [14]. That is the realization of a glucose prediction algorithm that, equipped in a modern CGM system, can be able to make direct predictions without a long period of data collection and training on the patient itself. In [14], we have demonstrated the possibility of designing an effective multi-patient data-driven blood glucose prediction model able to predict the future glucose level values of a new patient. This model allows the realization of a robust and pre-trained system. In this manuscript, we want to investigate how our LSTM solution works on the individual patient. Proved its potentialities on individual patients, as future works, we plan to apply some techniques, such as Transfer Learning [15], to personalize the multi-patients neural network according to glycaemia behaviours of each patient to monitor.

The rest of the paper is organised as follows. Section 2 describes the proposed methodology used to forecast blood glucose level in short- and medium-term. Section 3 debates the experimental results. Finally, Section 4 provides concluding remarks.

2. METHODOLOGY

This section describes our methodology for the realization of a data-driven patient-specialized solution based on LSTM neural network to forecast blood glucose level in short- and medium-term. Thus, we detail the dataset, and we introduce the state-of-art neural architectures (i.e. [9] and [10]) that represent our benchmark to evaluate and compare the performance of our solution. Consequently, starting from the literature structures, we design and discussing our characterization processes. In detail, we describe how we selected and optimised the final architecture and related parameters of the prediction models to boost the performances and accuracy, preventing the risk of overfitting.

In Figure 1, we show the main phases of our study. In the so-called training phase, training samples first undergo a pre-processing step. The unfiltered data are used as input to build a prediction model. In the test phase, new unseen and unfiltered data are fed into the trained models to obtain the final predictions.

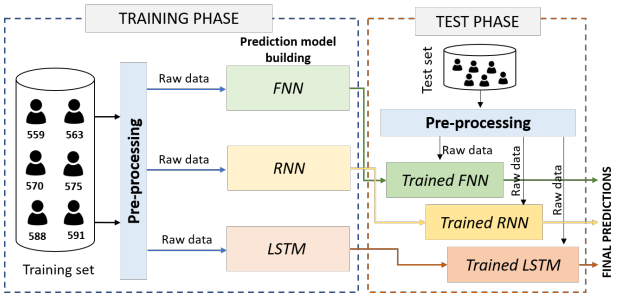


Fig. 1. Methodology - Main phases of the study.

2.1. Dataset and data pre-processing

To evaluate performance and accuracy, all the models are trained and validated using the same dataset: OhioT1DM [16]. This dataset consists of blood glucose level values sampled every 5 minutes, for about two months of observation. Data refer to six Type-I diabetic patients (i.e. two men and four women), with age ranging between 40 and 60 years. The sensor used to sample the blood glucose level is the Medtronic Enlite CGM, combined with a Medtronic 530G insulin pump. The dataset is divided into two parts: around 80% of the data for the training phase while the remaining 20% for the test and validation phase. However, following a preliminary phase check of the dataset, we noticed that patient measurements present some gaps. These faults may be due to temporary malfunctions or to the maintenance activities of the sensors. Such temporal lacks can lead to a wrong evaluation by the algorithms with a consequent loss of accuracy of the prediction. Consequently, we performed a pre-processing of the data by interpolating and re-sampling linearly.

Finally, we organize the data as follows:

$$train_X = \begin{bmatrix} G_0 & G_1 & G_2 & \dots & G_{27} & G_{28} & G_{29} \\ G_{30} & G_{31} & G_{32} & \dots & G_{57} & G_{58} & G_{59} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix} \quad (1)$$

$$train_Y = \begin{bmatrix} G_3 & G_4 & G_5 & \dots & G_{30} & G_{31} & G_{32} \\ G_{33} & G_{34} & G_{35} & \dots & G_{60} & G_{61} & G_{62} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix} \quad (2)$$

where G represents a value of the dataset. The $train_X$ matrix contains the input values of the neural network. It is a matrix composed of 30 columns, i.e. the current measurement plus the 29 previous measurements. Instead, the $train_Y$ matrix contains the output values. The two matrices have the same size, however, $train_Y$ is shifted by as many values as the prediction time horizon of interest. In the given example, the $train_Y$ is shifted by a number of samples equal to the

forecast horizon, i.e. 15 minutes, then 3 samples, one every 5 minutes (sampling rate). Furthermore, to avoid over-training, samples already used by the training network are no longer used, which is why the measurements go from G_0 to G_{30} (see $train_X$). Test matrices have the same structure as those for training. To avoid erroneous results, the values within these files are never used for the training phase, but are given as input to the network only when the performance of the final architecture is to be evaluated.

2.2. Prediction models

In the following, we introduce our LSTM solution. Moreover, we present the identified and re-implemented state-of-art models, a FNN [9] and a RNN [10], respectively, as reported by authors. Indeed, starting from the original structure, we try to optimize these models by modifying some hyperparameters, to obtain even more robust and high-performance models according to our dataset. The original FNN was composed of 30 inputs, two hidden layers of 10 and 5 units and an output layer [9]. The structure of the RNN is similar but with two hidden layers of 20 and 13 neurons, respectively [10]. For all the models, we selected the best architecture with the best configuration. Thus, starting from the literature, we evaluated different architectures' setups in terms of number of inputs, number of units in the hidden layer and iterations through a trial-and-error approach.

As far as FNN and RNN are concerned, our re-implemented solutions always consist of a single hidden layer. Indeed, we have managed to simplify the networks structures from a computational point of view, achieving similar or better performance.

2.2.1. Long Short-Term Memory Neural Network

LSTM represents the powerful evolution of the classic RNN. This neural model, particularly suitable for time series, is designed to overcome the limitations of the RNN. Indeed, RNN architectures suffer the instability of long-term predictions due to either vanishing or exploding gradient problems [17]. Often these problems arise during the training of deep structures, when the error gradients are propagated back in time to the initial layer, going through continuous matrix multiplications. To overcome this limitation, the LSTM is designed as *cells* where specific gating functions manage the memory, maintaining or erasing information at each time step. The ability to remove or add information to the cell state is regulated by gates (i.e. input, output, and forget) consisting of sigmoidal activation functions coupled with pointwise multipliers. Each gate modulates how much of the corresponding signal should be let through. In this way, LSTMs are able to remember values that are passed through gates all in 1 state, regardless of how deep the network is.

To find the best neural structure, we investigated different structural combinations with a trial-and-error approach. We found that the best structure consists of 30 inputs, a layer composed of 50 cells and an output layer. The number of optimal iterations is 2000, and the learning rate is 0.001, with the Adaptive Moment Estimation (Adam) as optimization algorithm [18].

2.2.2. Feed-forward Neural Network

This type of models represents the simplest type of ANN in which information propagation is mono-directional. In other words, the information moves from the input nodes to the output nodes, through the hidden layers [19]. Generally, this model is characterized by a fully-connected structure. It is also famous for its low computational costs [19]. Then, starting from the model architecture presented in [9], we performed different structure combinations. Then, we found the best structure that is composed of 30 inputs, 100 neurons in the hidden layer and only one neuron for the output layer. We exploited the sigmoid as activation function and the back-propagation as optimization, as generally recommended in literature [9]. Furthermore, we set the learning rate equal to 0.001, with an optimal number of iterations of 1500.

2.2.3. Recurrent Neural Network

The RNNs propagate the information in a bidirectional way. Indeed, they are composed of recurrent units sharing the same parameters, with loops allowing to propagate the information back to the same computational units. In this way, each computational step takes into account not only the current input but also the previous ones. As in the FNN case, starting from the model presented in [10], we re-implemented the structure by investigating different configurations. The best structure is configured as a follow: i) 30 input units, ii) 50 neurons in the hidden layer and iii) a single output. We exploited the sigmoid as activation function. As optimization method, we used Adam with a learning rate of 0.001 and a number of iterations set to 3000.

3. EXPERIMENTAL RESULTS

In this section, we discuss the experiments performed to assess the prediction models described in Section 2. We split our analysis in two main part for a thorough evaluation of the applied methods: i) *Analytical assessment:* we exploit a set of statistical indexes widely used in literature to validate the predictions from a regression analysis point of view; ii) *Clinical assessment:* we exploit metrics specifically designed to validate the clinical outcome of blood glucose measurements to validate the predictions from a clinical point of view. To evaluate the performance of our LSTM, our results are compared with the results obtained by literature techniques based on a FNN [9] and a RNN [10], as described in Section 2.2.

3.1. Analytical assessment

To evaluate the prediction accuracy of the models, we exploited a set of metrics that are often used by blood glucose level predictions literature [20]. These metrics allow quantifying the similarities between predicted and observed time-series trough descriptive statistics and regression analysis. The *RMSE - Root Mean Square Error* is the standard deviation of the difference between the predicted and the observed values. It represents the prediction error-index that is the most used in literature. The R^2 - *Coefficient of Determination* is defined as square of the correlation (R) between predicted and observed values. Thus, it ranges from 0 (absence of correlation) to 1 (complete correlation).

Table 1. Prediction performance indicators for FNN.

		Index	Patient ID					
			559	563	570	575	588	591
FNN	30 min	RMSE (mg/dl)	22.22	20.49	18.53	25.13	20.52	23.79
		r^2	0.88	0.79	0.92	0.80	0.80	0.79
	45 min	RMSE (mg/dl)	28.08	31.49	24.74	31.68	27.00	30.29
		r^2	0.81	0.55	0.86	0.72	0.65	0.65
	60 min	RMSE (mg/dl)	33.14	34.59	31.81	40.57	33.70	35.63
		r^2	0.73	0.47	0.77	0.55	0.46	0.49
	90 min	RMSE (mg/dl)	43.43	40.25	44.16	51.59	41.71	42.85
		R^2	0.51	0.27	0.58	0.36	0.21	0.29

Table 2. Prediction performance indicators for RNN.

		Index	Patient ID					
			559	563	570	575	588	591
RNN	30 min	RMSE (mg/dl)	18.26	21.00	17.53	22.53	19.62	23.46
		r^2	0.92	0.78	0.93	0.84	0.80	0.79
	45 min	RMSE (mg/dl)	24.00	32.06	23.71	30.26	26.24	30.12
		r^2	0.86	0.54	0.87	0.74	0.65	0.65
	60 min	RMSE (mg/dl)	30.05	35.48	30.87	37.15	33.04	35.61
		r^2	0.77	0.45	0.78	0.62	0.46	0.49
	90 min	RMSE (mg/dl)	40.07	41.47	42.31	47.44	42.07	43.46
		r^2	0.56	0.23	0.61	0.45	0.20	0.27

Tables 1, 2 and 3 report all the experimental results obtained by performing the re-implemented FNN and RNN architectures and our proposed LSTM solution. In all three scenarios, we used raw data in training, validation and inference phase for individual patients (marked with an identification code 559, 563, 570 575, 588, 591 respectively). Moreover, the forecast time horizons of interest are 30, 45, 60 and 90 minutes, respectively. For each prediction time horizon, we report values of $RMSE$ and R^2. As these networks are patient-specialised, consequently, we carried out the analysis of the results per patient.

Analyzing the analytical indices for all the patients, we found that state-of-art RNN is more performant than the FNN in terms of prediction accuracy. This translates into smaller

Table 3. Prediction performance indicators for our LSTM solution.

		Index	Patient ID					
			559	563	570	575	588	591
LSTM Proposed	30 min	RMSE (mg/dl)	11.52	10.84	10.20	9.29	11.55	10.56
		r^2	0.93	0.95	0.97	0.97	0.92	0.95
	45 min	RMSE (mg/dl)	19.58	12.05	16.14	18.32	19.86	16.91
		r^2	0.90	0.94	0.94	0.90	0.79	0.87
	60 min	RMSE (mg/dl)	27.67	20.88	23.02	29.07	25.00	25.46
		r^2	0.80	0.85	0.88	0.76	0.68	0.71
	90 min	RMSE (mg/dl)	43.99	36.03	34.24	49.89	30.95	41.44
		r^2	0.52	0.63	0.74	0.36	0.54	0.29

values of $RMSE$ and R^2, as described in Table 1 and Table 2 respectively. However, this happens for all time horizons with the exception of the Patient 563. In this specific case, the indices for each prediction time horizon are slightly higher. This phenomenon occurs also for our proposed LSTM solution (see Table 3). Also, for the patient 591, we noted that, with the exception of a few individual time horizons (i.e. 60 an 90 minutes), our LSTM solution achieve slightly worse results. In our opinion, this is because the dataset at our disposal is extremely small for adequate training, especially for recurrent networks, which generally require a large enough amount of data to better understand their trends and non-linear relationships [21]. However, even with this limitation due to the dataset, our LSTM achieves better performance. For patients 559 and 588, the results are much better for each time horizon analyzed (i.e. from 30 min. to 90 min.). Instead, Patient 570 is the best case found. In fact, for each time horizon, we lower the RMSE value from 6 mg/dl to 8 mg/dl compared to RNN, which in turn performs better than FNN. Generally, we can attest that our solution based on a neural network type LSTM results the most promising architecture. In terms of prediction accuracy, compared to all the investigated time horizons, we found improvements at every time horizon under analysis.

Referring to the most recent literature, the maximum acceptable RMSE value ranges from 19.32 mg/dl to 24.83 mg/dl in 30 min. predictions [22]. Using these state-of-art limits as a benchmark, we can achieve good prediction accuracy up to 60 min. in the future for Patients 563 and 570. This limit can also be considered suitable for Patients 588 and 591, as the maximum deviation is about 0.6 mg/dl.

Figure 2 shows a plot of the RSME values obtained by all patients for our proposed LSTM solution. The horizontal continuous black lines represent the best and the worst limit for 30 min. predictions, as described in [22]. Taking as reference the most conservative limit (i.e. the lowest black line), we can affirm that our model allows us to achieve good results up to 45 min. for all patients and up to 60 min. ahead for Patients 588 and 591.

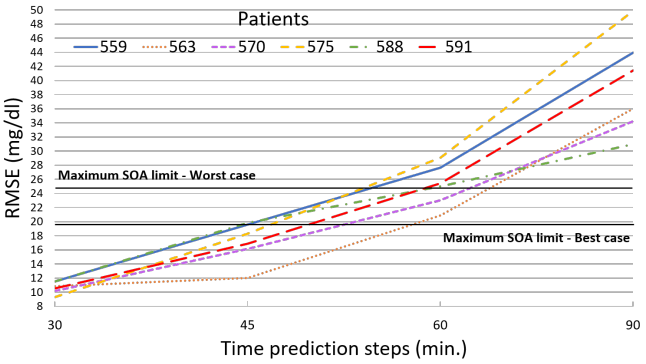


Fig. 2. LSTM predictions analysis.

3.2. Clinical assessment

The identified analytical metrics (see Section 3.1) are suitable to understand the performances and the prediction accuracy of models from a regression analysis point of view. However, these metrics are not appropriate to identify the most significant outliers, and they do not provide any information about the clinical impact of the prediction errors and their consequences on medical treatment decisions. Therefore, to give a more thorough picture of the models' performance, we integrated our assessment with Clarke Error Grid analysis (EGA) [23]. EGA represents one of the most widely accepted indexes for the analysis of clinical accuracy of blood glucose estimations. Indeed, it provides a clinical interpretation of the mapping between predicted and measured blood glucose levels, that can be represented in a scatter plot with five regions: *A-B* - values that, despite being outside 20% of the reference, do not lead to inadequate treatment of the patient; *C* - values leading to inappropriate treatment, but without dangerous consequences for the patient; *D-E* - values leading to potentially dangerous failure to detect hypoglycaemic or hyperglycaemic events. Zones A and B are fully clinical acceptable. D and E refer to the zones where prediction errors are mostly dangerous [24].

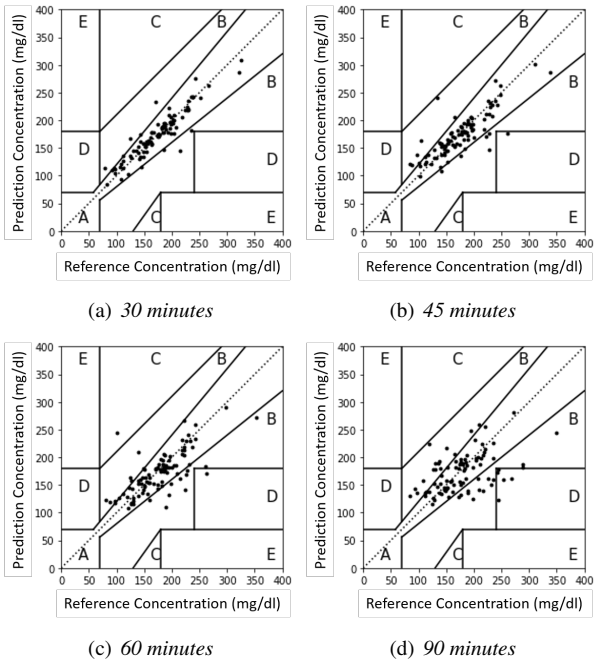
Figure 3 shows the EGA analysis performed for all time horizons (i.e. from 30 to 90 min.) for Patient 575. This represents the worst case we found, in terms of prediction accuracy, especially for the 60 and 90 min. prediction horizons. The EGA analysis confirms what was found analytically in the Section 3.1. Up to 45 minutes in the future, all values are concentrated in the zones fully clinical acceptable (i.e. A and B). Differently, for longer time horizons some values are placed in zone D, therefore already potentially dangerous for the patient.

4. CONCLUSIONS

CGMs sensors represent a valid and more advantageous alternative to self-monitoring for patients affected by Type-I diabetes. Through the diffusion of such devices, nowadays, a considerable amount of data are available. This asset allows the study and implementation of different and innovative models in the field, especially those based on neural networks techniques.

In this work, we addressed the problem of automated glucose level prediction leveraging patient-specialized CGMs data. In detail, we have designed, implemented and compared different specialized state-of-art models based on neural networks. Our specific aim is to find the best neural solution that best fits the specialization on the individual patient. Thus, starting with a deeply specialized dataset and some neural solution in literature, we identify the best-optimized structure, with the best prediction performance in terms of forecast accuracy.

In our future work, we will combine our patient-specialized data-driven system with our multi-patient data-driven methodology [14] by integrating run-time information. More specifically, we plan to perform a real-time fine-tuning of the model, leveraging the glucose level measurements of the patient that is currently using the system.



(a) *30 minutes* (b) *45 minutes*

(c) *60 minutes* (d) *90 minutes*

Fig. 3. Clarke Error Grids for specialized LSTM network for Patient 575.

5. REFERENCES

[1] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[2] Edoardo Patti, Maria Donatelli, Enrico Macii, and Andrea Acquaviva, "Iot software infrastructure for remote monitoring of patients with chronic metabolic disorders," in *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*. IEEE, 2018, pp. 311–317.

[3] Md Zia Uddin, Mohammed Mehedi Hassan, Ahmed Alsanad, and Claudio Savaglio, "A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare," *Information Fusion*, vol. 55, pp. 105–115, 2020.

[4] Chiara Dalla Man, Francesco Micheletto, Dayu Lv, Marc Breton, Boris Kovatchev, and Claudio Cobelli, "The uva/padova type 1 diabetes simulator: new features," *Journal of diabetes science and technology*, vol. 8, no. 1, pp. 26–34, 2014.

[5] Souradeep Dutta, Taisa Kushner, and Sriram Sankaranarayanan, "Robust data-driven control of artificial pancreas systems using neural networks," in *International Conference on Computational Methods in Systems Biology*. Springer, 2018, pp. 183–202.

[6] Silvia Oviedo, Josep Vehí, Remei Calm, and Joaquim Armengol, "A review of personalized blood glucose prediction strategies for t1dm patients," *International journal for numerical methods in biomedical engineering*, vol. 33, no. 6, pp. e2833, 2017.

[7] Jaouher Ben Ali, Takoua Hamdi, Nader Fnaiech, Véronique Di Costanzo, Farhat Fnaiech, and Jean-Marc Ginoux, "Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 4, pp. 828–840, 2018.

[8] Takoua Hamdi, Jaouher Ben Ali, Véronique Di Costanzo, Farhat Fnaiech, Eric Moreau, and Jean-Marc Ginoux, "Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 2, pp. 362–372, 2018.

[9] Carmen Pérez-Gandía, A Facchinetti, G Sparacino, C Cobelli, EJ Gómez, M Rigla, Alberto de Leiva, and ME Hernando, "Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring," *Diabetes technology & therapeutics*, vol. 12, no. 1, pp. 81–88, 2010.

[10] Fayrouz Allam, Zaki Nossai, Hesham Gomma, Ibrahim Ibrahim, and Mona Abdelsalam, "A recurrent neural network approach for predicting glucose concentration in type-1 diabetic patients," in *Engineering Applications of Neural Networks*, pp. 254–259. Springer, 2011.

[11] Hrushikesh N Mhaskar, Sergei V Pereverzyev, and Maria D van der Walt, "A deep learning approach to diabetic blood glucose prediction," *Frontiers in Applied Mathematics and Statistics*, vol. 3, pp. 14, 2017.

[12] Kezhi Li, John Daniels, Pau Herrero Viñas, Chengyuan Liu, and Pantelis Georgiou, "Convolutional recurrent neural networks for blood glucose prediction," *CoRR*, vol. abs/1807.03043, 2018.

[13] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[14] Alessandro Aliberti, Irene Pupillo, Stefano Terna, Enrico Macii, Santa Di Cataldo, Edoardo Patti, and Andrea Acquaviva, "A multi-patient data-driven approach to blood glucose prediction," *IEEE Access*, vol. 7, pp. 69311–69325, 2019.

[15] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 9, 2016.

[16] Cindy Marling and Razvan Bunescu, "The ohiot1dm dataset for blood glucose level prediction," in *The 3rd International Workshop on Knowledge Discovery in Healthcare Data, Stockholm, Sweden*, 2018.

[17] Felix Gers, Jrgen Schmidhuber, and Fred Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, pp. 2451–71, 10 2000.

[18] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[19] Weipeng Cao, Xizhao Wang, Zhong Ming, and Jinzhu Gao, "A review on neural networks with random weights," *Neurocomputing*, vol. 275, pp. 278–287, 2018.

[20] Christian A Gueymard, "A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 1024–1034, 2014.

[21] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.

[22] John Martinsson, Alexander Schliep, Björn Eliasson, and Olof Mogren, "Blood glucose prediction with variance estimation using recurrent neural networks," *Journal of Healthcare Informatics Research*, pp. 1–18, 2019.

[23] William L Clarke, "The original clarke error grid analysis (ega)," *Diabetes technology & therapeutics*, vol. 7, no. 5, pp. 776–779, 2005.

[24] Daniel J Cox, Linda A Gonder-Frederick, Boris P Kovatchev, Diana M Julian, and William L Clarke, "Understanding error grid analysis," 1997.