Increasing Patient Specificity of the Recurrent Neural Network Based Insulin Sensitivity Prediction by Transfer Learning

Bálint Szabó
Dept. of Cont. Eng. and Inf. Tech.
Fac. of Elec. Eng. and Inf.
Budapest University of Tech. and Eco.
Budapest, Hungary
bszabo@iit.bme.hu

Ákos Szlávecz
Dept. of Cont. Eng. and Inf. Tech.
Fac. of Elec. Eng. and Inf.
Budapest University of Tech. and Eco.
Budapest, Hungary
szlavecz@iit.bme.hu

Katalin Kovács
Department of Informatics
Fac. of Mech. Eng., Inf. and Elect. Eng.
Széchenyi István University
Győr, Hungary
kovacsk@sze.hu

Béla Paláncz
Department of Geodesy and Surveying
Faculty of Civil Engineering
Budapest University of Tech. and Eco.
Budapest, Hungary
palancz.bela@epito.bme.hu

Geoffrey Chase
Department of Mechanical Engineering
University of Canterbury
Canterbury, New Zealand
geoff.chase@canterbury.ac.nz

Balázs Benyó
Dept. of Cont. Eng. and Inf. Tech.
Fac. of Elec. Eng. and Inf.
Budapest University of Tech. and Eco.
Budapest, Hungary
bbenyo@iit.bme.hu

2022 IEEE 26th International Conference on Intelligent Engineering Systems (INES) | 978-1-6654-9209-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/INES56734.2022.9922645

***Abstract*—Insulin therapy is a frequently applied treatment in intensive care to normalize the patient's blood glucose level increased by stress-induced hyperglycaemia. This therapy is generally referred to as Tight Glycaemic Control (TGC). The STAR (Stochastic-TARgeted) protocol is a TGC which uses the patient's insulin sensitivity (SI) as a key parameter to describe the patient's actual state. Prediction of the future patient's state, i.e. prediction of the patient's future SI value, is a crucial step of the protocol currently implemented by using the so-called Intensive Care INsulin Glucose (ICING) model of the human glucose-insulin system and an associated stochastic model. In our previous studies, we have shown that the Recurrent Neural Network (RNN) models are efficient alternative methods of SI prediction. In this paper, we suggest applying the so-called transfer learning technique to further enhance the accuracy of the SI prediction by using the SI history of the current patient. The paper presents the proposed methodology for applying transfer learning in SI prediction and the evaluation of the method's accuracy by comparing the outcomes with the currently applied solution. Insilico validation using real patients' data is involved in this validation.**

***Index Terms*—machine learning, artificial intelligence, mixture density network, deep neural network, insulin sensitivity, tight glycaemic control, intensive care, STAR protocol, validation, in-silico validation**

I. Introduction

Stress-induced hyperglycaemia is a frequent complication in the intensive therapy [1], [2] increasing the mortality and morbidity of the patients.

Controlling the blood glucose (BG) level of these hyperglycaemic patients into the so-called normoglycaemic range shows definite clinical benefits [3]–[6].

The STAR (Stochastic-TARgeted) TGC protocol is the most widely applied clinical therapy used to control the BG of the hyperglycaemic patients in the intensive care [7]. STAR uses a clinically validated physiological model, called Intensive Control Insulin-Nutrition-Glucose (ICING) to describe the glucose-insulin dynamics, and a population-based stochastic model to manage patient-specific metabolic variability [8].

STAR uses the patient-specific insulin sensitivity (SI) [9] in the method selecting optimal treatment. This method uses simulations with different treatment parameters and calculate the blood glucose (BG) levels at the end of these simulated treatments. From the results the protocol can recommend the optimal treatment parameters [10].

The treatment selection process [11] requires the prediction of a confidence interval on the future value of SI that comes from a conditional distribution. In the 2D case this distribution is made up from $\{SI(t);SI(t+1)\}$ data points [12]. The dimension refers to the dimension of the data points. We call the prediction higher dimensional if the data points in the distribution has more than 2 dimension.

The STAR protocol uses a confidence interval that contains the $SI(t+1)$ value with 90% probability in the prediction. To calculate this interval, the conditional distribution is approximated with normal distribution. With this approximation the 5th and the 95th percentile can be used as the border values of the interval. It has 90% probability that the $SI(t+1)$ will be between these two points, moreover the width of the interval is minimal.

Recently new artificial intelligence, especially neural network based models were created [13] with the aim of replacing the currently used prediction method. There are studies that analysed the effects of involving additional parameters into the prediction [14], [15]. They showed clear benefits by develop-

ing models on the so called 3D SI distributions. These models can easily handle the involvement of additional prediction parameter that makes possible to create even higher dimensional models and make the treatment more personalized.

In this paper the results of the in-silico validation [16] are presented when the clinical treatment is simulated on virtual patients that were created from the historical treatment data. In the next chapter the used artificial intelligence models and methods are presented together with the patient data set applied. In the third chapter the details of the results are introduced followed by the discussion in chapter four. The relevant results are summarized in chapter five, Conclusions.

II. Methods and Data

A. Patient Selection and SI Data Set Used

The patient's dataset under STAR treatment is collected in several studies. The training data used in this study was collected between June 2016 and August 2019 and filtered by the following excluding rules:

- patients treated less than 10 hours by STAR;
- sections of treatments where the higher border of the BG target band was above 9 mmol/L;
- sections of treatments where lower border of the BG target band was above 6 mmol/L.

The dataset consists of 797 SI trajectory, widely variated in length.

For the training the trajectories were split into 20 element long chunks. For the normalization the z-score normalization was used. This transforms the datapoints by the mean and variance of their distribution as they were zero and one.

The dataset was separated for training and test sets on a per-trajectory basis. 10% of the dataset was used for testing and from the remaining training set also 10% was used for validation to avoid overfitting during the training. The short trajectories on which the specific model can not be applied were also filtered out.

B. RNN

The recurrent neural networks have great potential of forecasting and processing time series in multiple application fields [17], such as natural language modelling [18], speech recognition [19] and image captioning [20]. In the biomedical field [21] it showed great potential in image segmentation [22], medical treatment optimization [23] and asynchronous event detection during mechanical-ventilation [24].

The base concept of RNNs is that during the model's training phase they extract the hidden state (see Figure 1) transition model encoded in the training dataset. Our solution uses the Gated Recurrent Unit [25] which shows better performance in several applications than the widely used Long Short Term Memory (LSTM) RNN architecture [26].

The proposed models use a sequence of historical SI values to predict the distribution of the future SI. During the training the number of inputs is fixed. However, due to the feedback loop, in the prediction the model can use arbitrary long SI sequences. This property makes a significant difference compared to the state of art methods of SI prediction. The current techniques used in practice are based on autoregressive kernel models. These methods do not scale well with the number of input parameters and datapoints used for the prediction. The prediction usually uses only the SI(t) value as a first-order Markov chain. Our RNN based method can extract information from arbitrary long SI series from which we expect more accurate prediction as more values become available during the treatment of the patients. The output is calculated not only from the input but also from the inner state. Therefore it is more expressive than the first-order Markov chain. Also, as a neural network based method, it scales well with the number of datapoints which only affects the offline training cost.

The model outputs are the mean and the variance of the predicted SI distribution. Assuming a normal distribution the 5th and 95th percentile define a 90% confidence interval required by the protocol. Due to this special output configuration, a special loss function have to be used as well. This loss function can be resulted from the negative log likelihood of the normal distribution.

The models are trained on 20 long SI time series chunks. This long sequences tends to face overfitting. Thus, early stopping was applied with the help of a 10% validation set separated from the training data.

C. Transfer learning

Transfer learning is a machine learning method in which a pre-trained model is trained on another dataset from a similar problem class. Transfer learning can fine-tune the model for a specific, new task and reduce the time and computation cost of the training as it is kickstarted from a general model.

Transfer learning was used to create patient specific models by training the RNN model on the trajectory of the patient under treatment. At the beginning of the treatment the specific model can not be used because there is no sufficient data to train on. Therefore, the initial treatment phase where the general model could be applied only was excluded from the evaluation.

The specific model can be created when there are at least one more measurement in the trajectory than the recurrence length of the model. After that, the specific model can be recreated as new BG measurements are gathered. The creation of a specific model means that a copy of the general model is created and trained with the data of the given patient's trajectory for some epochs. The number of measurements between recreation and the number of training epochs is hyperparameters that can significantly affect the method's performance.

In this study, we recreated the model when additional measurements were taken in the size of the recurrence length. This means that the specific model was recreated and trained at every 20 new measurements for 10 epochs. We expect that the transfer learning moves the model to predict the specific distribution of the given patient, potentially leading to more accurate prediction. We observed that the variance of the whole

dataset is larger than one can occur in the scope of one patient's trajectory. Our motivation is to exploit this observed property and narrow the confidence intervals by using transfer learning.

D. Metrics

To compare the newly created methods with each other and with the currently used method we need metrics that are close to the application requirements. The success rate (SR) metric captures the 90% confidence requirement. It is calculated by dividing the number of predictions where the future SI was in the predicted interval by the total number of datapoints:

$$m_{\mathrm{SR}} = \frac{N_{\mathrm{in}}}{N} \tag{1}$$

where m_{SR} is the success rate, N_{in} is the number of datapoints that are in the predicted interval, N is the number of datapoints, l_i and h_i are the endpoints of the i-th confidence interval and t_i is the i-th datapoint. The greater SR is preferred.

The narrower predicted intervals assumed to be beneficial for the protocol because they gives more space to optimize the treatment parameters. With the Interval Ration (IR) metric we want to capture this beneficial property of the newly created model compared to another prediction method. The metric is calculated as:

$$m_{\mathrm{IR}} = \frac{1}{N}\sum_{i=0}^{N}\frac{\mathrm{h}_i - \mathrm{l}_i}{r_i}, \tag{2}$$

where m_{IR} is the interval ratio, N is the number of datapoints, l_i and h_i are the endpoints of the i-th confidence interval, and r_i is the width of predicted interval calculated by the other prediction method as a basis of comparison. The lesser IR is preferred.

These two properties are in a trade-off relation. We created a single metric that helps us evaluate the relation. It is called i-score (IS) and calculated as:

$$m_{\mathrm{IS}} = \frac{\mathrm{icdf}\left(0.5 + \frac{m_{\mathrm{SR}}}{2}\right)}{\mathrm{icdf}(0.95)\cdot m_{\mathrm{IR}}} \tag{3}$$

where m_{SR} is the success rate, m_{IR} is the average interval ratio and icdf is the inverse cumulative distribution function of the standard normal distribution. The higher i-score is preferred.

All metrics can be calculated on a per datapoint basis and on a per-patient basis. More about these metrics can be read in [27].

III. Results

In the evaluation we wanted information on a global and on a per-patient basis. Averaging the results on the whole dataset gives us an image of the performance of the models in the long while on a per-patient basis it helps us to identify unique properties of the models based on a specific trajectory and also to detect anomalies and resolve them during the model development.

Table I shows the numerical, per-datapoint results of the newly developed specific method compared to the general RNN based method. The individual intervals has a distribution

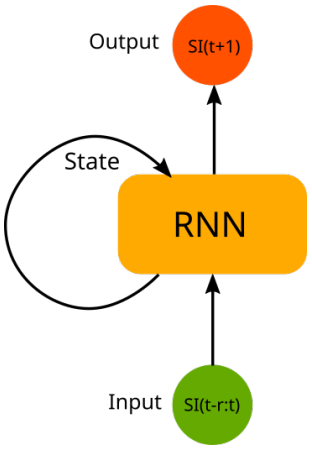

Fig. 1. Schematic figure of the RNN architecture.

TABLE I
STATISTICAL RESULTS OF THE GENERAL AND SPECIFIC MODEL.

	General	Specific
Success rate	0.9034	0.8812
Interval ratio	1.0	0.9111

which can be seen in Figure 2 as a histogram with marked mean.

In the per-patient evaluation success rate also creates a distribution as it can be evaluated for each patient. This success rate histograms of the two models can be seen in Figure 6. The per-patient average interval ratio histogram can be seen in Figure 3.

Because the success rate and interval ratio are in a trade-off relation the per-patient results were visualized in a scatter plot where the x axis was the success rate and the y axis is the interval ratio. In this plot we can also draw the i-score = 1 curve. In this case the datapoints under the curve represents the treatments where the success rate was improved or the interval ratio was reduced as much that they belong to a better prediction performance. This plot can be seen in Figure 5.

Figure 4 shows how the trajectory length affects the success rate, as the values of the trajectory creates the transfer learning dataset for the specific model. We also visualized some actual trajectory to compare the two methods. On these plots, the specific model starts later as it gets enough information for the first transfer learning cycle. For both methods the confidence intervals were visualized as well. These trajectories can be seen in Figure 7-10.

It is important to note that these results were produced on a test trajectory set which was excluded from the training set and even from the validation set. So these results represent the practical case with new, unseen patients of the models.

IV. Discussion

The numerical results from Table I show that the newly developed method does not improved on the success rate metrics on a global scale. The success rate decreased to 88%

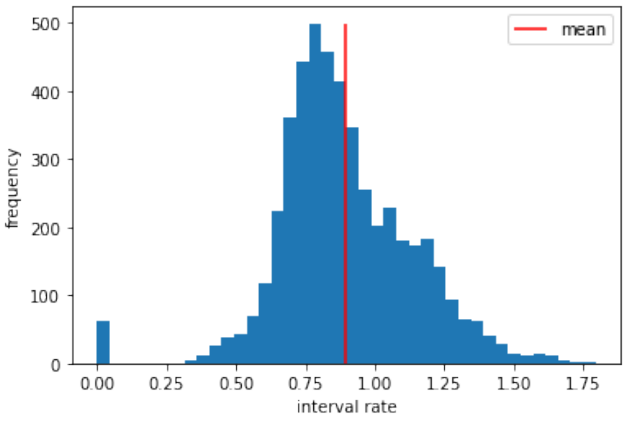

Fig. 2. Histogram of the interval ratio per datapoint. The red line is the mean of the distribution.

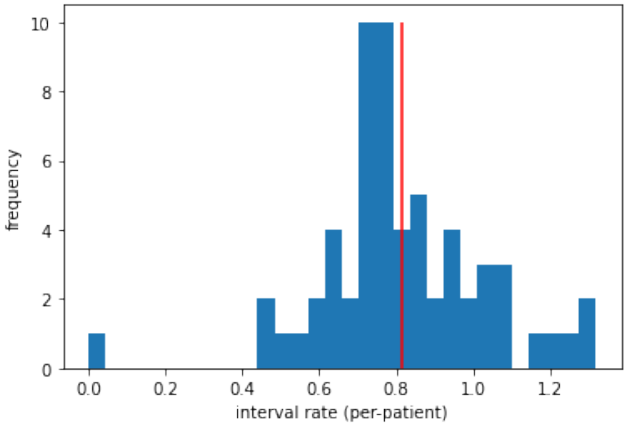

Fig. 3. Histogram of the interval ratio per patient. The red line is the mean of the distribution.

which can be tolerated. Figure 4 shows that the lowest per-patient success rates comes from the shortest treatment length. This tells us that the new method has a requirement on the data quantity after which it can produce as good predictions as the general model. Therefore the application of new method is preferred in the long, protracted treatments. The high outliers of this figures (where the success rate is 1.0) comes from the cases when the treatment length is close to the recurrence size of the model.

The per-datapoint and per-patient interval ratio histograms (Figure 2 and 3) show that the mean interval ratio of the new method under one in both case. This means that in a regular case the new method reduces the interval widths which is beneficial for the treatment optimization. In Figure 2 it can be seen that there is significant amount of datapoint where the intervals were widened. On the per patient basis it is not that significant. The per-patient success rate histogram of the specific model became wider. It has lower values than the histogram of the general model. We assume that the reason behind these results is the interval reducing manner of the

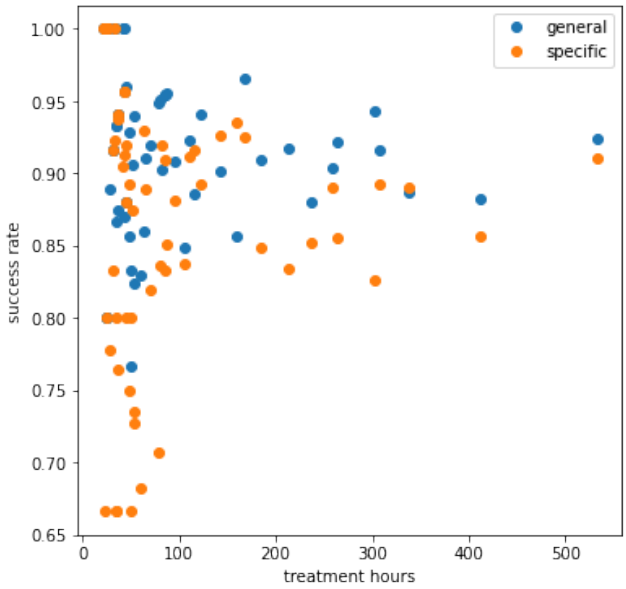

Fig. 4. Scatter plot of the success rate based on the length of the treatment.

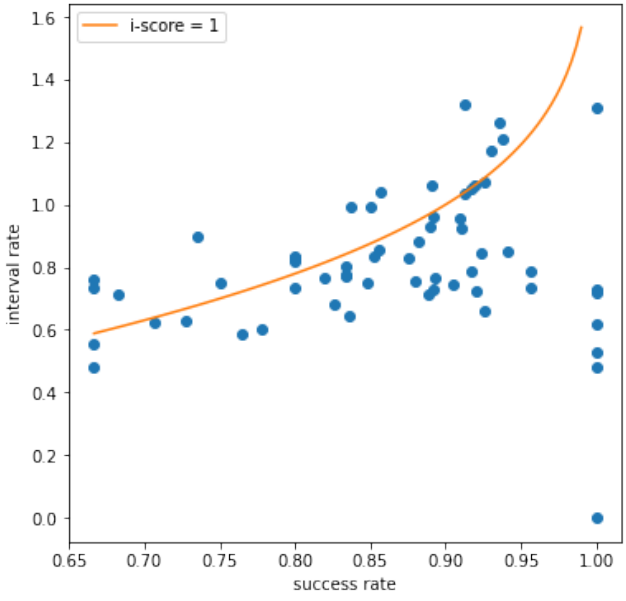

Fig. 5. Scatter plot of the interval ratio based on success rate. The orange curve represent the 1 i-score value. Under the curve the trajectory has a greater i-score than 1.

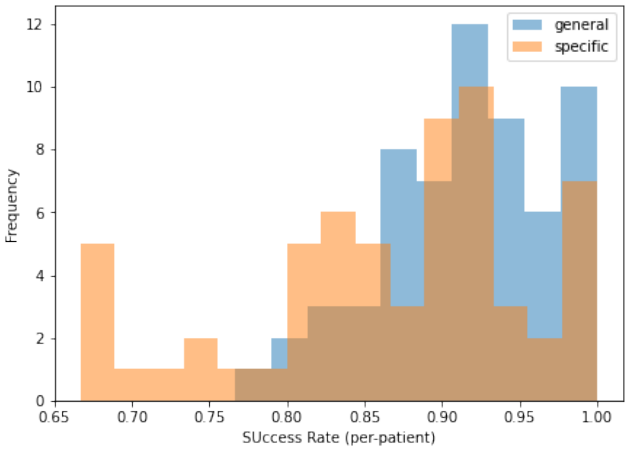

Fig. 6. Histogram of the per-patient success rate for the general (blue) and the specific (orange) models.

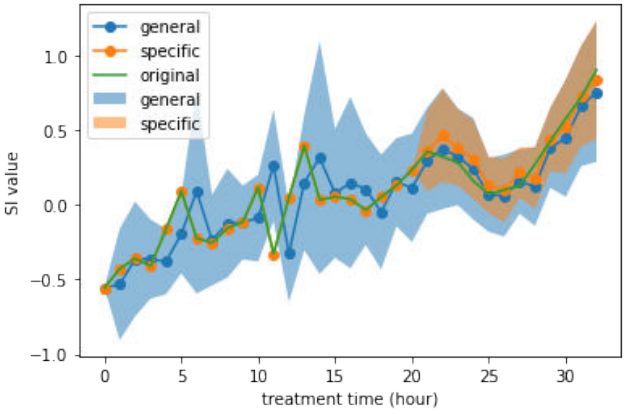

Fig. 7. Example prediction trajectory. The green curve is the original trajectory, blue is the general model and orange is the specific method. The confidence intervals are coloured as well.

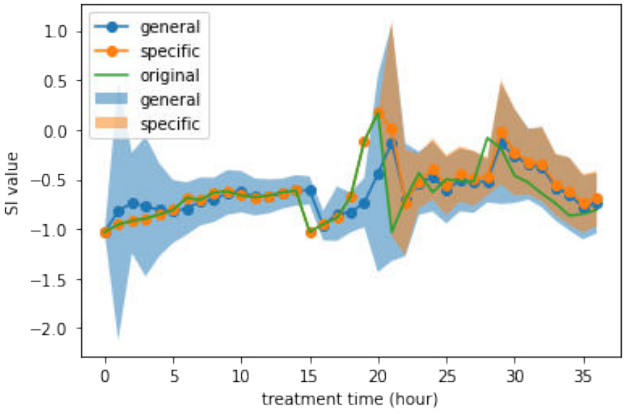

Fig. 8. Example prediction trajectory. The green curve is the original trajectory, blue is the general model and orange is the specific method. The confidence intervals are coloured as well.

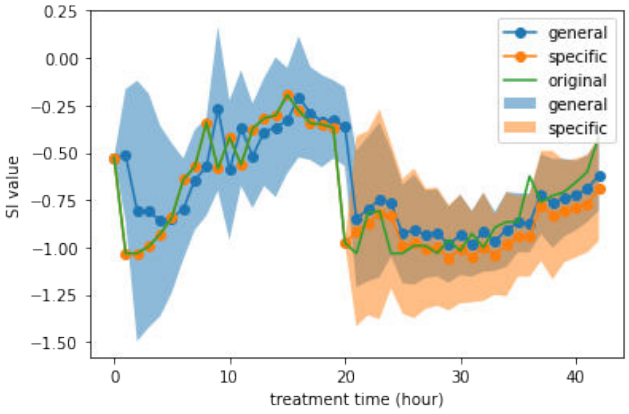

Fig. 9. Example prediction trajectory. The green curve is the original trajectory, blue is the general model and orange is the specific method. The confidence intervals are coloured as well.

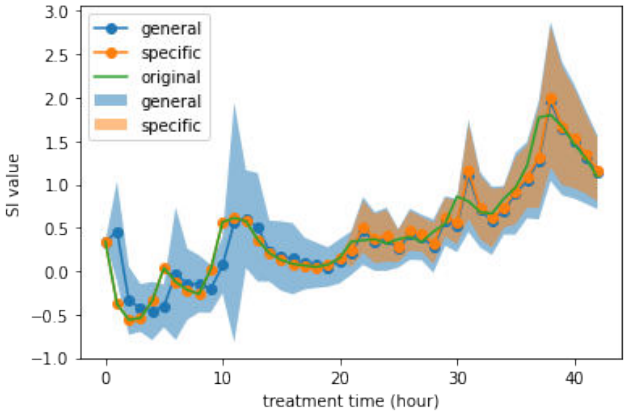

Fig. 10. Example prediction trajectory. The green curve is the original trajectory, blue is the general model and orange is the specific method. The confidence intervals are coloured as well.

new specific model.

Figure 5 shows how the per-patient success rate and average interval ratio relates to each other. The i-score=1 curve is displayed on the figure as well. Under this curve the cases have larger i-score than one which means that they improved on the success rate and interval ratio trade-off. There are numerous cases under this curve. Therefore, we can say that there are patients whose treatment benefited from the specific model.

From the evaluated trajectories we concluded that the specific model can significantly reduce the interval width compared to the general model. An interesting property is that the intervals are often reduced only from the lower endpoint. There is also an exception example in Figure 9. Here we can see that the intervals of the specific model is wider than the ones of the general models. It is not clear what combination of properties led to this anomaly but one can notice that the initial transfer learning phase of the specific model connects to a great drop on the SI trajectory.

Overall, it can be said that the novel method can enhance the

treatment in some special cases but the general applicability of the model did not reach our expectations. However, the development of the novel method deepened our understanding on the SI trajectory. In the further research the diversity of the SI trajectories should be distinguished as intra-patient and inter-patient diversity. We should analyse how the SI can change in the scope of the treatment of the patient (intra-patient) and how the patients can differ in their SI trajectory.

V. Conclusions

In this research a new neural network based method was presented. The method applies transfer learning on a general recurrent network model to increase patient specificity by train it with a specific insulin sensitivity trajectory.

We found this new method beneficial in several cases as it can reduce the width of the predicted confidence intervals. The prediction performance of the novel method greatly depends on the length of the treatment. Therefore it is suggested to be used in the longer treatments.

As a further step of the research we plan to analyse the variances of the patient specific SI trajectories and how they relate to the global datapoint variance.

Acknowledgment

The research was supported by the Hungarian Scientific Research Fund (OTKA), Grant K137995, by the National Research, Development, and Innovation Fund of Hungary under Grant TKP2021-EGA-02, and by the European Union's H2020 MSCA-RISE DCPM (#872488) grant.

References

[1] K. C. McCowen, A. Malhotra, and B. R. Bistrian, "Stress-induced hyperglycemia," *Critical care clinics*, vol. 17, no. 1, pp. 107–124, 2001.

[2] N. A. Ali, J. M. O'Brien Jr, K. Dungan, G. Phillips, C. B. Marsh, S. Lemeshow, A. F. Connors Jr, and J.-C. Preiser, "Glucose variability and mortality in patients with sepsis," *Critical care medicine*, vol. 36, no. 8, p. 2316, 2008.

[3] G. van den Berghe, P. Wouters, F. Weekers, C. Verwaest, F. Bruyninckx, M. Schetz, D. Vlasselaers, P. Ferdinande, P. Lauwers, and R. Bouillon, "Intensive insulin therapy in critically ill patients," *New England journal of medicine*, vol. 345, pp. 1359–1367, 11 2001.

[4] J. S. Krinsley, "Effect of an intensive glucose management protocol on the mortality of critically ill adult patients," in *Mayo Clinic Proceedings*, vol. 79, no. 8. Elsevier, 2004, pp. 992–1000.

[5] R. et al, "Intensive insulin protocol improves glucose control and is associated with a reduction in intensive care unit mortality," *Journal of the American College of Surgeons*, vol. 204, no. 5, pp. 1048–1054, 2007.

[6] J. G. Chase, C. G. Pretty, L. Pfeifer, G. M. Shaw, J.-C. Preiser, A. J. Le Compte, J. Lin, D. Hewett, K. T. Moorhead, and T. Desaive, "Organ failure and tight glycemic control in the SPRINT study," *Crit Care*, vol. 14, no. 4, p. R154, 2010.

[7] K. Stewart, H. Tomlinson, F. Thomas, J. Homlok, S. Noémi, A. Illyés, G. Shaw, B. Benyó, and J. Chase, "Safety, efficacy and clinical generalization of the star protocol: a retrospective analysis," *Annals of Intensive Care*, vol. 6, p. 24, 12 2016.

[8] A. Evans, A. Le Compte, C.-S. Tan, L. Ward, J. Steel, C. G. Pretty, S. Penning, F. Suhaimi, G. M. Shaw, T. Desaive *et al.*, "Stochastic targeted (STAR) glycemic control: design, safety, and performance," *Journal of diabetes sci. and tech.*, vol. 6, no. 1, pp. 102–115, 2012.

[9] B. Benyó, "In-silico analysis of stochastic modelling of human blood glucose regulatory system," in *2019 IEEE 23rd International Conference on Intelligent Engineering Systems (INES)*, 2019, pp. 000 229–000 234.

[10] B. I. Benyó, B. Paláncz, Á. J. Szlávecz, B. Szabó, Y. Anane, K. Kovács, and J. G. Chase, "Artificial intelligence based insulin sensitivity prediction for personalized glycaemic control in intensive care," *IFAC PAPERSONLINE*, 2020.

[11] J. G. Chase, B. I. Benyó, and T. Desaive, "Glycemic control in the intensive care unit: A control systems perspective," *ANNUAL REVIEWS IN CONTROL*, vol. 48, pp. 359–368, 2019.

[12] B. Szabó, B. Paláncz, Á. J. Szlávecz, K. Kovács, and B. I. Benyó, "Artificial intelligence based insulin sensitivity prediction for personalized glycaemic control in intensive care," in *Orvosi Informatika 2020 – A XXXIII. Neumann Kollokvium konferencia kiadványa*, 2020, pp. 96–101.

[13] B. Szabó, C. Geoffrey, B. Paláncz, Á. J. Szlávecz, Y. Anane, K. Kovács, and B. I. Benyó, "Deep neural network based methods for predicting human insulin sensitivity in thight glycaemic control," in *Proceedings of the Workshop on the Advances of Information Technology 2020*, 2020, pp. 39–46.

[14] V. Uyttendaele, J. Knopp, K. Stewart, T. Desaive, B. Benyó, N. Szabó-Némedi, A. Illyés, G. Shaw, and J. Chase, "A 3d insulin sensitivity prediction model enables more patient-specific prediction and model-based glycaemic control," *Biomedical Signal Processing and Control*, vol. 46, pp. 192–200, 09 2018.

[15] V. Uyttendaele, J. L. Knopp, S. Davidson, T. Desaive, B. Benyo, G. M. Shaw, and J. G. Chase, "3D kernel-density stochastic model for more personalized glycaemic control: development and in-silico validation," *BioMedical Engineering OnLine*, vol. 18, no. 1, p. 102, 2019.

[16] B. Szabó, Á. J. Szlávecz, B. Paláncz, C. Geoffrey, and B. I. Benyó, "In-silico simulation based evaluation of insulin prediction method for personalized medical treatment," in *Proceedings of the Workshop on the Advances of Information Technology (WAIT) 2021*, 2021, pp. 145–155.

[17] B. Palancz, "Stochastic forecasting of time series via rnn using lst memory layer," 03 2022.

[18] W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Computer Speech and Language*, vol. 30, no. 1, pp. 61–98, 2015.

[19] G. Saon, Z. Tüske, D. Bolanos, and B. Kingsbury, "Advancing rnn transducer technology for speech recognition," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 5654–5658.

[20] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*. PMLR, 2015, pp. 2048–2057.

[21] E. Atoyebi, S. Adedayo, N. Idusuyi, and A. Olorunnisola, "A review of artificial neural networks for biomedical applications: Trends and prospects," 09 2018.

[22] M. Rezaei, H. Yang, and C. Meinel, "Recurrent generative adversarial network for learning imbalanced medical image semantic segmentation," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15 329–15 348, 2020.

[23] H. Le, T. Tran, and S. Venkatesh, "Dual control memory augmented neural networks for treatment recommendations," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2018, pp. 273–284.

[24] L. Zhang, K. Mao, K. Duan, S. Fang, Y. Lu, Q. Gong, F. Lu, Y. Jiang, L. Jiang, W. Fang, X. Zhou, J. Wang, L. Fang, H. Ge, and Q. Pan, "Detection of patient-ventilator asynchrony from mechanical ventilation waveforms using a two-layer long short-term memory neural network," *Computers in Biology and Medicine*, vol. 120, p. 103721, 2020.

[25] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014.

[26] S. Yang, X. Yu, and Y. Zhou, "Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example," in *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, 2020, pp. 98–101.

[27] B. Szabó, Á. J. Szlávecz, B. Paláncz, J. G. Chase, and B. I. Benyó, "Comparison of gated recurrent unit vs. mixture density network in insulin sensitivity prediction."