PREDICTING ADVERSE EVENTS FOR PATIENTS WITH TYPE-1 DIABETES VIA SELF-SUPERVISED LEARNING

Xinzhe Zheng *Sijie Ji* *Chenshu Wu*

Department of Computer Science, The University of Hong Kong
{xzzheng, sijieji, chenshu}@cs.hku.hk

ABSTRACT

Predicting blood glucose levels is fundamental for precise primary care of type-1 diabetes (T1D) patients. However, it is challenging to predict glucose levels accurately, not to mention the early alarm of adverse events (hyperglycemia and hypoglycemia), namely the minority class. In this paper, we propose BG-BERT, a novel self-supervised learning framework for blood glucose level prediction. In particular, BG-BERT incorporates masked autoencoder to capture rich contextual information of blood glucose records for accurate prediction. More specifically, SMOTE data augmentation and shrinkage loss are employed to effectively handle adverse events without discrimination. We evaluate BG-BERT on two benchmark datasets against two state-of-the-art baseline models. The experimental results highlight the significant improvements achieved by BG-BERT in glucose level prediction accuracy (measured by RMSE) and sensitivity to adverse events, with average lifting ratios of 9.5% and 44.9%, respectively.

***Index Terms*—** type-1 diabetes, glucose prediction, self-supervised learning, hyperglycemia, hypoglycemia

1. INTRODUCTION

Type-1 diabetes (T1D) is a prevalent chronic condition worldwide, and it is not curable. The only therapy is to maintain blood glucose levels by injecting an appropriate amount of insulin [1]. Improper insulin dosages often lead to the occurrence of adverse events: notably hyperglycemia (blood glucose levels less than 70mg/dL) and extreme hypoglycemia (blood glucose levels greater than 250mg/dL). Hypoglycemia can have an impact on an individual's electrocardiogram and may even result in a loss of consciousness [2]. Hyperglycemia has the potential to cause harm to the ischemic brain and increase the risk of stroke [3]. Therefore, managing blood glucose levels is critical for individuals with T1D who require long-term self-administration of insulin.

Certain advanced medical devices incorporate Continuous Glucose Monitoring (CGM) capabilities to autonomously regulate insulin administration and mitigate the occurrence of adverse events [4]. Recently, blood glucose level prediction has achieved sensational progress with the development of deep learning [5, 6]. In particular, with a series of past blood glucose records (e.g., blood glucose levels, insulin doses, carbs intake), the neural network model, especially recurrent neural networks (RNNs), can predict blood glucose levels on the horizon. For example, DRTF [7] utilizes a fully connected block structure with an RNN to forecast gradually in each block, and wins the 2020 Blood Glucose Level Prediction Challenge [8]. Based on the architecture of DRTF, MT-NB-L [9] adds an auxiliary branch to learn blood glucose forecasting and the probability of hypoglycemia events in a multi-task learning manner, and also achieves promising results. However, despite the remarkable advancements in accuracy achieved by the latest deep learning-based solutions for blood glucose prediction, accurately predicting glucose levels within the adverse event range remains a challenge. This is primarily attributed to the insufficient modeling of contextual information in blood glucose records and the limited availability of blood glucose record data within the adverse event range. These limitations pose obstacles to the development of reliable adverse event prediction models.

To address the aforementioned challenge, we propose BG-BERT, a novel blood glucose prediction framework. Building on top of BERT [10], BG-BERT harnesses the power of masked autoencoder mechanism [11] during the self-supervised learning phase to effectively model contextual information of blood glucose monitoring data. This enables the model to understand the interplay between glucose levels and adjacent data, capturing both fluctuating trends and the influence of observed data on future readings. Furthermore, BG-BERT incorporates advanced techniques to specifically address the issue of limited adverse event-associated blood glucose monitoring data. In particular, the data augmentation method known as Synthetic Minority Over-sampling Technique (SMOTE) [12] is employed to create synthesized data for minority adverse events before feeding the data into the network. A shrinkage loss function [13] is adopted during the training process to guide the model to predict glucose levels within adverse events without bias or discrimination. To assess the performance of BG-BERT, we compare it against two prevailing baseline models, namely DRTF [7] and MT-NB-L [9]. The evaluation results substantiate that

BG-BERT achieves significantly higher accuracy in predicting blood glucose levels within the range of adverse events than the baseline models. BG-BERT is open-sourced here: `https://github.com/aiot-lab/BG-BERT`.

2. PROBLEM STATEMENT

The subject of this paper is to predict adverse events for T1D. We do not consider it as a classification problem but try to predict the exact blood glucose levels on the horizon. The predicted glucose levels can both carry the variation trend and the occurrence of these adverse events, which would benefit the insulin injection or pumps in clinical applications. The glucose levels to be predicted are determined by historical glucose monitoring data, including glucose levels, carbs intake, bolus insulin dose, basal insulin rate, etc. Given N previous glucose monitoring data, denoted as $\mathbf{X}_{1:N} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$, the prediction model should be able to precisely estimate the following blood glucose levels within the target horizon (TH). In this paper, we use $\mathbf{g}_{N+1:N+TH} = [g_{N+1}, ..., g_{N+TH}]$ to represent the predicted blood glucose levels. To evaluate the identification capability for adverse events, we labeled all data points (g_i) with blood glucose levels below 70mg/dL as hypoglycemic events and data points above 250mg/dL as severe hyperglycemic events.

3. METHODOLOGY

Inspired by the success of BERT [10] for NLP tasks, which is particularly good at extracting and modeling contextual information from temporal data, we propose BG-BERT that builds on top of BERT but is tailored for blood glucose prediction. The BG-BERT comprises two phases: self-supervised learning and glucose level predicting, as illustrated in Fig.1. The masked self-supervised learning phase enables the encoder to develop a more profound comprehension of the factors influencing glucose levels and their potential future implications. The acquired representations are used to enhance the subsequent glucose level prediction phase.

First, to address imbalanced data and the lack of adverse event samples, we employ the data augmentation technique known as SMOTE. In particular, the synthetic instances are generated by combining a specific sample from adverse events with one of its k nearest neighbors. The calculation for generating the synthetic object is as follows:

$$\mathbf{X}_s = \mathbf{X}_i + r \cdot (\mathbf{X}_b - \mathbf{X}_i) \tag{1}$$

where $\mathbf{X}_s$ is the synthetic instance, $\mathbf{X}_i$ is one sample from adverse events, and $\mathbf{X}_b$ denotes the neighbor sample of $\mathbf{X}_i$. We apply SMOTE only on training set for data augmentation.

After that, span-mask [14] is implemented to mask the monitoring data $\mathbf{x}$ at selected time stamps. The span-mask selects the mask length in geometric distribution, making the mask positions more continuous. Hence, BG-BERT could learn the temporal relationship in glucose monitoring data more effectively. The masked data is then fed into the encoder to learn the deep representations $\mathbf{R}$ such that:

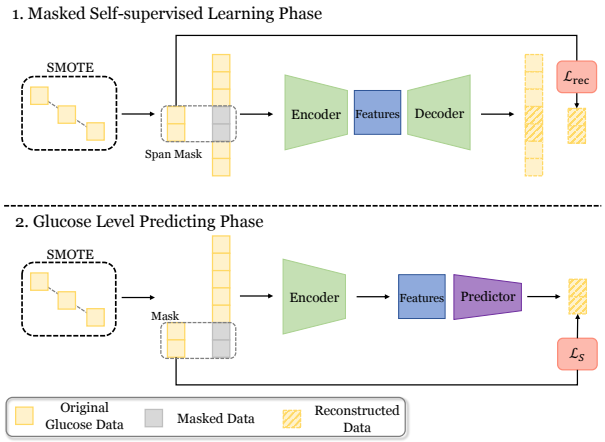

Fig. 1. System overview of BG-BERT.

$$\mathbf{R} = \mathrm{Enc}(\mathbf{X}_m^u + \mathrm{PE}(\mathbf{X}_m^u)) \tag{2}$$

where Enc represents the encoder network, $\mathbf{X}_m^u$ denotes the masked uniform glucose monitoring data, and PE is the position encoding to make full use of the order information. Then, the representations will be used to reconstruct the original values of the masked glucose data given the equation of $\hat{\mathbf{X}}^u = \mathrm{Dec}(\mathbf{R})$, where $\hat{\mathbf{X}}^u$ denotes the reconstructed glucose data, and Dec is the decoder network. We only calculate the reconstruction loss $\mathcal{L}_{rec}$ for the masked samples following the training strategy of BERT using mean squared error (MSE):

$$\mathcal{L}_{rec} = \mathrm{MSE}(\mathrm{Mask}(\mathbf{X}^u), \mathrm{Mask}(\hat{\mathbf{X}}^u)). \tag{3}$$

For encoder architecture, we use pre-norm residual unit [15] to connect multi-head attention layers which can further capture rich contextual information. In particular, we implement add-normed multilayer perceptions (MLP) for decoder network design.

In the glucose level predicting phase, we mask the target horizon in the glucose sequence and feed it into the encoder with fixed parameters to generate representations. A predictor, composed of LSTM, CNN, and MLP modules, is designed to forecast glucose levels based on the learned features $\mathbf{R}$. The LSTM module handles the sequential information in the representations, while the CNN module combines historical and estimated information. Lastly, the MLP module determines the expected changes in glucose levels. The estimated blood glucose levels $\hat{\mathbf{g}}$ is obtained with $\hat{\mathbf{g}} = \mathrm{Pred}(\mathbf{R})$. We don't fine-tune the encoder but train the predictor only because fine-tuning the whole network would achieve almost the

same performance as using representation exclusively. To enhance the predictor's focus on adverse events, we draw inspiration from the focal loss function [16] used for imbalanced data classification. We amplify large loss values and diminish small loss values during back-propagation, implementing the concept of focal loss for regression tasks. We utilize a customized shrinkage loss function [13] for glucose level prediction, the formula is given as:

$$\mathcal{L}_s = \frac{||\hat{\mathbf{g}} - \mathbf{g}||^2}{1 + \exp(a \cdot (c - ||\hat{\mathbf{g}} - \mathbf{g}||^2))} \tag{4}$$

where a and c are hyper-parameters of shrinkage loss.

4. EXPERIMENT SETUP

4.1. Benchmark Dataset

The evaluation is conducted on two benchmark datasets: OhioT1DM [8] and Diatrend [17]. OhioT1DM contains 8 weeks of data for 12 patients with T1D, making it the most commonly used dataset for glucose level prediction. Only 3.4% of the readings in OhioT1DM are hypoglycemia events, while 8.2% are hyperglycemia events. Diatrend is the largest open-source glucose monitoring dataset, with 27,561 days of continuous glucose monitoring data from 54 T1D patients. In Diatrend, the occurrence rate of hypoglycemia is 1.35%, while hyperglycemia events account for 10.24% of the total recordings. Both datasets are used with agreements from corresponding affiliations. We partition each dataset as 64:16:20 for training, validation, and testing respectively, following previous works. We set up two sets of experiments, where the combinations of N and TH are (2-hour, 30-min) and (4-hour, 60-min) respectively, following previous works [7, 18]. As the continuous glucose monitor records data every five minutes, the number of N and TH corresponds to (24, 6) and (48, 12). These settings also have clinical implications considering the interstitial fluid blood glucose lag time [19] and insulin onset time [20].

4.2. Baseline Models

We take two state-of-the-art glucose prediction models as baselines for comparison. DRTF won the 2020 Blood Glucose Level Prediction Challenge [8], and MT-NB-L presents a multitask learning approach to predict glucose level and adverse events simultaneously. Both baseline models use a supervised learning strategy.

4.3. Evaluation Metrics

We evaluate the performance from four perspectives.

- **Root Mean Squared Error (RMSE)**: RMSE measures the average distance between predicted glucose levels $\hat{g}_i$ and the golden standard data g_i, which is given as $\sqrt{\sum_i (\hat{g}_i - g_i)^2}$.
- **Temporal Gain (TG)**: It indicates the amount of average time gained for early detection of a potential adverse event using the model. Corresponding to a sampling time Δt, a total of N samples, and a L-step ahead prediction horizon, the calculation procedure is given as:

$$\text{delay} = \underset{i \in [0,L]}{\text{argmin}} \frac{1}{N-L} \sum_{k=1}^{N-L} \left(\hat{g}(k+i) - g(k)\right)^2 \tag{5}$$

$$\text{TG} = (L - \text{delay}) \cdot \Delta t \tag{6}$$

- **Sensitivity**: Sensitivity indicates the ability of a test to correctly identify adverse events.
- **Clarke Error Grid** [21]: Clarke error grid is used to quantify the risk of using measured or predicted values of blood glucose for glucose management of diabetic patients.

5. EVALUATION

5.1. Overall Performance Evaluation

We evaluate BG-BERT's performance compared to baseline models and analyze the impact of its components through an ablation study. The evaluation is done on the OhioT1DM and Diatrend datasets, considering both 30-minute and 60-minute time horizons. The summarized results are presented in Tab.1.

BG-BERT demonstrates significant improvements compared to the baseline models. For the 30-minute horizon prediction, BG-BERT achieves RMSE results of 14.02 mg/dL and 14.85 mg/dL on the OhioT1DM and Diatrend datasets, respectively, surpassing DRTF and MT-NB-L. The self-supervised learning strategy employed by BG-BERT enhances its understanding of glucose variations, leading to improved average early detection time (TG). Through the utilization of SMOTE data augmentation and a shrinkage loss function, BG-BERT exhibits an average increase of 46.4% in hypoglycemia sensitivity compared to the baseline models, and a slight improvement in hyperglycemia sensitivity.

BG-BERT demonstrates consistent excellence in predicting glucose levels over a 60-minute period. It excels in reducing prediction errors (RMSE) and accurately detecting hyperglycemia. On average, BG-BERT improves RMSE by 9.5% and significantly enhances hypoglycemia sensitivity compared to other models.

To further understand the contribution of different components in BG-BERT, an ablation study is conducted. By removing data augmentation (w/o aug), we observe a performance loss of 4.33% compared to BG-BERT. Additionally, when shrinkage loss is removed (w/o $\mathcal{L}_s$), the model cannot handle the accurate prediction of the minority class. These findings indicate that both SMOTE augmentation and shrinkage loss contribute to the performance gains of BG-BERT.

Overall, our evaluation results demonstrate the effectiveness of leveraging BERT-based architectures for blood glucose prediction.

Table 1. Overall performance on two benchmark datasets: (**Red**: the best performance among all algorithms. **Hype Sen** and **Hypo Sen** indicate the detection sensitivity of hyperglycemia and hypoglycemia respectively.)

Horizon	30 mins								60 mins							
Dataset	OhioT1DM				Diatrend				OhioT1DM				Diatrend			
Metric	RMSE (mg/dL)	TG (mins)	Hype Sen(%)	Hypo Sen(%)	RMSE (mg/dL)	TG (mins)	Hype Sen(%)	Hypo Sen(%)	RMSE (mg/dL)	TG (mins)	Hype Sen(%)	Hypo Sen(%)	RMSE (mg/dL)	TG (mins)	Hype Sen(%)	Hypo Sen(%)
DRTF	18.21	15.54	80.67	53.58	15.23	15.07	80.28	39.12	28.36	28.7	68.89	23.38	26.42	28.17	58.93	16.74
MT-NB-L	21.50	14.74	61.36	34.05	19.80	13.88	75.16	39.89	34.15	26.34	41.65	25.40	25.59	28.72	62.32	12.07
w/o aug	14.38	16.19	81.49	64.41	15.01	16.27	79.53	56.69	23.64	31.14	66.62	48.46	25.22	31.39	62.65	32.49
w/o $\mathcal{L}_s$	**13.92**	15.72	81.16	70.75	15.13	16.16	80.75	57.42	**22.61**	30.04	68.31	30.21	25.03	31.03	64.13	31.01
BG-BERT	14.02	**16.56**	**82.54**	**73.24**	**14.85**	**16.47**	**81.34**	**62.27**	23.67	**31.16**	**69.24**	**54.12**	**24.95**	**31.45**	**64.53**	**40.1**

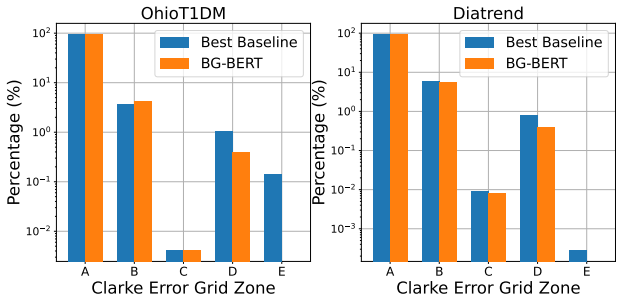

Fig. 2. Clarke error grid analysis on two benchmark datasets. (A: medically accurate result, B: medically acceptable, C: unnecessary treatment, D: failure to detect a dangerous condition, E: mistaking adverse events.)

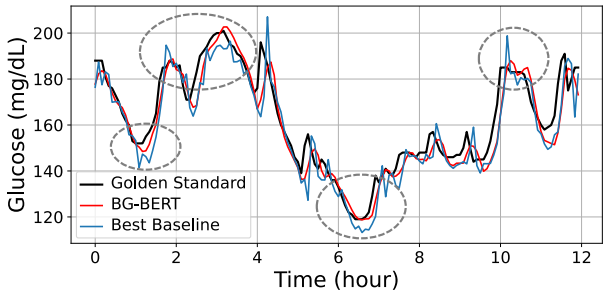

Fig. 3. Visualization of half-day glucose prediction. The gray circles highlight the better performance on turning points.

5.2. Clarke Error Grid Analysis

The present study employs the Clarke error grid analysis to assess clinical acceptability by comparing predicted values with golden standard values. Results are summarized in Fig. 2. The analysis reveals that BG-BERT outperforms the baseline model on both datasets in regions A and B, indicating its capability of precisely predicting glucose levels. When evaluating regions of C, D, and E, which determine the unacceptable results, the proportion of BG-BERT's prediction results in these regions is consistently lower compared to the baseline model. This disparity is particularly pronounced in region E, where BG-BERT's prediction results account for 0% of the total. The findings indicate that BG-BERT not only achieves improved prediction accuracy but also boasts a reduced false alarm rate.

5.3. Glucose Prediction Visualization

Fig. 3 displays the half-day glucose prediction results of BG-BERT in comparison to the best baseline model (DRTF). It is evident that BG-BERT accurately anticipates changes in blood glucose over time, revealing the effectiveness of mastering contextual information through self-supervised learning. Moreover, the reduced fluctuation highlights the prediction robustness of our proposed method.

6. CONCLUSION

This paper introduces BG-BERT, a novel self-supervised learning framework designed to predict blood glucose levels in patients with T1D. By incorporating a span-mask mechanism, BG-BERT enhances its understanding of the relationships between historical and future blood glucose levels, thereby improving prediction accuracy. To address the issue of imbalanced adverse event samples during the prediction stage, this study introduces two techniques: SMOTE data augmentation and a shrinkage loss function. The application of these techniques yields significant improvements in adverse event prediction, particularly in detecting hypoglycemia alarms, which observe an average progress of 46% on two benchmark datasets. We believe that BG-BERT has the potential to contribute to the research community and introduce new methodologies that advance blood glucose monitoring for the clinical care of T1D patients.

7. ACKNOWLEDGMENTS

This work is supported in part by NSFC under grant No. 62222216 and Hong Kong RGC ECS under grant No. 27204522 and RIF under grant No. R5060-19.

8. REFERENCES

[1] Anastasia Katsarou, Soffia Gudbjörnsdottir, Araz Rawshani, Dana Dabelea, Ezio Bonifacio, Barbara J Anderson, Laura M Jacobsen, Desmond A Schatz, and Åke Lernmark, "Type 1 diabetes mellitus," *Nature reviews Disease primers*, vol. 3, no. 1, pp. 1–17, 2017.

[2] Brian M Frier, Guntram Schernthaner, and Simon R Heller, "Hypoglycemia and cardiovascular risks," *Diabetes care*, vol. 34, no. Suppl 2, pp. S132, 2011.

[3] Perttu J Lindsberg and Risto O Roine, "Hyperglycemia in acute stroke," *Stroke*, vol. 35, no. 2, pp. 363–364, 2004.

[4] Hood Thabit and Roman Hovorka, "Coming of age: the artificial pancreas for type 1 diabetes," *Diabetologia*, vol. 59, no. 9, pp. 1795–1805, 2016.

[5] Jinyu Xie and Qian Wang, "Benchmarking machine learning algorithms on blood glucose prediction for type i diabetes in comparison with classical time-series models," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3101–3124, 2020.

[6] Yuting Xing, Hangting Ye, Xiaoyu Zhang, Wei Cao, Shun Zheng, Jiang Bian, and Yike Guo, "A continuous glucose monitoring measurements forecasting approach via sporadic blood glucose monitoring," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 860–863.

[7] Harry Rubin-Falcone, Ian Fox, and Jenna Wiens, "Deep residual time-series forecasting: Application to blood glucose prediction.," in *KDH@ ECAI*, 2020, pp. 105–109.

[8] Cindy Marling and Razvan Bunescu, "The ohiot1dm dataset for blood glucose level prediction: Update 2020," in *CEUR workshop proceedings*. NIH Public Access, 2020, vol. 2675, p. 71.

[9] Mu Yang, Darpit Dave, Madhav Erraguntla, Gerard L Cote, and Ricardo Gutierrez-Osuna, "Joint hypoglycemia prediction and glucose forecasting via deep multi-task learning," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 1136–1140.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[11] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16000–16009.

[12] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[13] Xiankai Lu, Chao Ma, Bingbing Ni, Xiaokang Yang, Ian Reid, and Ming-Hsuan Yang, "Deep regression tracking with shrinkage loss," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 353–369.

[14] Huatao Xu, Pengfei Zhou, Rui Tan, Mo Li, and Guobin Shen, "Limu-bert: Unleashing the potential of unlabeled data for imu sensing applications," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 220–233.

[15] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V Le, and Ruslan Salakhutdinov, "Transformer-xl: Attentive language models beyond a fixed-length context," *arXiv preprint arXiv:1901.02860*, 2019.

[16] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[17] Temiloluwa Prioleau, Abigail Bartolome, Richard Comi, and Catherine Stanger, "Diatrend: A dataset from advanced diabetes technology to enable development of novel analytic solutions," *arXiv preprint arXiv:2304.06506*, 2023.

[18] Taiyu Zhu, Lei Kuang, John Daniels, Pau Herrero, Kezhi Li, and Pantelis Georgiou, "Iomt-enabled real-time blood glucose prediction with deep learning and edge computing," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 3706–3719, 2022.

[19] Eda Cengiz and William V Tamborlane, "A tale of two compartments: interstitial versus blood glucose monitoring," *Diabetes technology & therapeutics*, vol. 11, no. S1, pp. S–11, 2009.

[20] Werner Regittnig, Martina Urschitz, Barbara Lehki, Michael Wolf, Harald Kojzar, Julia K Mader, Martin Ellmerer, and Thomas R Pieber, "Insulin bolus administration in insulin pump therapy: effect of bolus delivery speed on insulin absorption from subcutaneous tissue," *Diabetes Technology & Therapeutics*, vol. 21, no. 1, pp. 44–50, 2019.

[21] William L. Clarke, "The original clarke error grid analysis (ega)," *Diabetes Technology & Therapeutics*, vol. 7, no. 5, pp. 776–779, 2005, PMID: 16241881.