Supplementary Information: Deep Learning for Blood Glucose Prediction: Investigating Reproducibility and Factors Affecting Differential Performance

Supplementary Table 1: Model and training details for six selected deep learning methods in prior literature for blood glucose prediction.

Model Structure and Training Policy				
Method	**Model Type**	**Scaling**	**Train test split**	**Activation**
Martinsson 2019[1]	LSTM	0.01	80-20	exp
Li 2020[2]	Dilated CNN	0.01	80-20	relu
van Doorn 2021[3]	LSTM	0.01	80-20	relu
Deng 2021[4]	CNN	0.01	80-20 + fine tune	relu
Rabby 2021[5]	LSTM	0.01	80-20	relu + exp
Lee 2023[6]	Transformer	LayerNorm	80-20	relu

Network Architecture	
Method	**Network Layers**
Martinsson 2019[1]	1 LSTM (256) + 2 Dense (512, 256)
Li 2020[2]	5 Conv (32, 32, 64, 64, 128)
van Doorn 2021[3]	2 LSTM (32, 16) + 1 Dense (8)
Deng 2021[4]	4 blocks with 2×1DConv + 2 Dense (10, 10)
Rabby 2021[5]	2 LSTM (512 + 128 + 1)
Lee 2023[6]	4 Multi-head attention encoders + Linear(512)

Training Configuration				
Method	**Optimization**	**Loss Function**	**Regularization**	**Epochs**[1]
Martinsson 2019[1]	Adam	NLL	DropOut(0.2, 0.3)	10000 (200)
Li 2020[2]	Adam	RMSE	Early Stopping	100
van Doorn 2021[3]	Adam	MSE	DropOut(0.1)	10000 (100)
Deng 2021[4]	Adam	RMSE	L2 regularization	80
Rabby 2021[5]	Adam	MSE	DropOut (0.2)	6000 (128)
Lee 2023[6]	Adam	RMSE	Dropout(0.2)	200 (25)

Method	**Learning Rate**	**Batch Size**
Martinsson 2019[1]	10^{-3}	1024
Li 2020[2]	10^{-4}	32
van Doorn 2021[3]	10^{-4}	1024
Deng 2021[4]	10^{-3}	64
Rabby 2021[5]	10^{-3}	128
Lee 2023[6]	10^{-3}, weight_decay=10^{-4}	64

[1]Maximum training epochs (early stopping patience)

Supplementary Table 2: RMSE (mg/dL) per participant for predicting blood glucose 30 mins ahead using a zero-order hold baseline.

Public T1D Datasets	Mean	Std Dev	Max	Min
OhioT1DM [7]	23.21	2.89	28.40	18.81
DiaTrend [8]	28.72	5.28	42.50	20.78
T1DEXI [9]	24.14	5.06	40.32	13.47

Supplementary Table 3: Model performance on different datasets

Method	Mean	Std Dev	Max	Min
OhioT1DM: 12 subjects				
Martinsson 2019[1]	18.86	2.13	21.63	16.04
Li 2020[2]	19.31	2.38	22.63	15.56
van Doorn 2021[3]	19.20	2.14	22.12	16.36
Deng 2021[4]	19.34	2.30	22.35	16.10
Rabby 2021[5]	19.78	2.37	22.74	16.25
Lee 2023[6]	19.78	2.45	23.15	16.67
Aggregate Stats	**19.38**	**2.30**	**23.15**	**15.56**
DiaTrend: 53 subjects				
Martinsson 2019[1]	23.85	3.91	34.18	17.52
Li 2020[2]	26.20	6.11	46.83	17.81
van Doorn 2021[3]	24.08	3.89	34.18	17.56
Deng 2021[4]	24.24	3.91	34.52	17.76
Rabby 2021[5]	25.24	4.59	35.96	18.26
Lee 2023[6]	25.46	4.56	36.80	17.83
Aggregate Stats	**24.85**	**4.50**	**46.83**	**17.52**
T1DEXI: 63 subjects				
Martinsson 2019[1]	19.40	3.79	32.98	10.81
Li 2020[2]	20.92	4.33	33.03	10.81
van Doorn 2021[3]	19.55	3.71	33.01	11.06
Deng 2021[4]	19.80	3.73	32.63	11.13
Rabby 2021[5]	22.25	4.59	36.08	11.62
Lee 2023[6]	20.31	3.96	34.20	11.84
Aggregate Stats	**20.37**	**4.02**	**36.08**	**10.81**

Supplementary Table 4: T-test results comparing model performance across datasets

Method	OhioT1DM vs DiaTrend	OhioT1DM vs T1DEXI
Martinsson 2019[1]	$t = -4.26$, $p = 6.9 \times 10^{-5}$ ***	$t = -0.48$, $p = 0.635$
Li 2020[2]	$t = -3.82$, $p = 9.0 \times 10^{-5}$ ***	$t = -1.25$, $p = 0.217$
van Doorn 2021[3]	$t = -4.18$, $p = 9.7 \times 10^{-5}$ ***	$t = -0.32$, $p = 0.752$
Deng 2021[4]	$t = -4.16$, $p = 9.8 \times 10^{-5}$ ***	$t = -0.41$, $p = 0.681$
Rabby 2021[5]	$t = -3.99$, $p = 1.76 \times 10^{-4}$ ***	$t = -1.84$, $p = 0.070$
Lee 2023[6]	$t = -4.16$, $p = 3.03 \times 10^{-4}$ ***	$t = -0.45$, $p = 0.657$

Supplementary Table 5: Correlation between individual diabetes management indicators and RMSE (mg/dL) for predicting blood glucose 30 minutes ahead.

Method	Avg. BG vs RMSE $(r,\ p)$	TIR vs RMSE $(r,\ p)$	GV vs RMSE $(r,\ p)$	HbA1c vs RMSE $(r,\ p)$
Martinsson 2019[1]	$(0.55,\ 1.8{\times}10^{-11})$	$(-0.58,\ 6.5{\times}10^{-13})$	$(0.51,\ 1.0{\times}10^{-9})$	$(0.47,\ 1.7{\times}10^{-7})$
Li 2020[2]	$(0.67,\ 1.1{\times}10^{-17})$	$(-0.65,\ 6.3{\times}10^{-17})$	$(0.39,\ 6.5{\times}10^{-6})$	$(0.45,\ 7.9{\times}10^{-7})$
van Doorn 2021[3]	$(0.56,\ 5.5{\times}10^{-12})$	$(-0.60,\ 9.5{\times}10^{-14})$	$(0.52,\ 3.2{\times}10^{-10})$	$(0.48,\ 7.2{\times}10^{-8})$
Deng 2021[4]	$(0.56,\ 6.8{\times}10^{-12})$	$(-0.59,\ 1.7{\times}10^{-13})$	$(0.51,\ 8.4{\times}10^{-10})$	$(0.48,\ 6.3{\times}10^{-8})$
Rabby 2021[5]	$(0.57,\ 3.1{\times}10^{-12})$	$(-0.56,\ 4.3{\times}10^{-12})$	$(0.38,\ 8.7{\times}10^{-6})$	$(0.34,\ 2.6{\times}10^{-4})$
Lee 2023[6]	$(0.64,\ 5.2{\times}10^{-16})$	$(-0.66,\ 1.7{\times}10^{-17})$	$(0.50,\ 2.2{\times}10^{-9})$	$(0.52,\ 4.0{\times}10^{-9})$

Supplementary Table 6: Difference in 30-min prediction RMSE between male and female subgroups and corresponding t-test results.

	30-min BG Pred. RMSE (mg/dL)				
Method	**Male** (Mean ± SD)	**Female** (Mean ± SD)	**Difference**	**t-statistic**	**p-value**
Martinsson 2019[1]	19.86 ± 3.94	21.84 ± 4.33	1.99	-2.49	0.014*
Li 2020[2]	21.36 ± 4.71	23.73 ± 5.96	2.37	-2.24	0.027*
van Doorn 2021[3]	20.10 ± 3.91	22.03 ± 4.31	1.93	-2.42	0.017*
Deng 2021[4]	20.26 ± 3.89	22.24 ± 4.32	1.98	-2.50	0.014*
Rabby 2021[5]	21.54 ± 4.54	24.12 ± 4.60	2.58	-2.96	0.004**
Lee 2023[6]	20.79 ± 4.14	23.17 ± 4.93	2.38	-2.68	0.008**

Supplementary Table 7: Comparing blood glucose prediction models trained with CGM plus other lifestyle factors (i.e. carbohydrate input, insulin data, and step count) versus CGM only as input features. These methods were trained and evaluated using the OhioT1DM dataset.

	Mean RMSE (mg/dL)	
Method	Using CGM + Others	Using CGM only
Rabby 2021 [5]	20.35	19.81
Li 2020 [2]	20.34	18.93

References

[1] Martinsson, J., Schliep, A., Eliasson, B., Mogren, O.: Blood glucose prediction with variance estimation using recurrent neural networks. Journal of healthcare informatics research **4**, 1–18 (2019) https://doi.org/10.1007/S41666-019-00059-Y

[2] Li, K., Liu, C., Zhu, T., Herrero, P., Georgiou, P.: Glunet: A deep learning framework for accurate glucose forecasting. IEEE journal of biomedical and health informatics **24**, 414–423 (2020) https://doi.org/10.1109/JBHI.2019.2931842

[3] van Doorn, W.P.T.M., Foreman, Y.D., Schaper, N.C., Savelberg, H.H.C.M., Koster, A., Kallen, C.J.H., Wesselius, A., Schram, M.T., Henry, R.M.A., Dagnelie, P.C., de Galan, B.E., Bekers, O., Stehouwer, C.D.A., Meex, S.J.R., Brouwers, M.C.G.J.: Machine learning-based glucose prediction with use of continuous glucose and physical activity monitoring data: The maastricht study. PLOS ONE **16**, 0253125 (2021) https://doi.org/10.1371/JOURNAL.PONE.0253125

[4] Deng, Y., Lu, L., Aponte, L., Angelidi, A.M., Novak, V., Karniadakis, G.E., Mantzoros, C.S.: Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. npj Digital Medicine 2021 4:1 **4**, 1–13 (2021) https://doi.org/10.1038/s41746-021-00480-x

[5] Rabby, M.F., Tu, Y., Hossen, M.I., Lee, I., Maida, A.S., Hei, X.: Stacked lstm based deep recurrent neural network with kalman smoothing for blood glucose prediction. BMC Medical Informatics and Decision Making **21**, 1–15 (2021) https://doi.org/10.1186/S12911-021-01462-5/TABLES/7

[6] Lee, S.-M., Kim, D.-Y., Woo, J.: Glucose transformer: Forecasting glucose level and events of hyperglycemia and hypoglycemia. IEEE Journal of Biomedical and Health Informatics **27**(3), 1600–1611 (2023)

[7] Marling, C., Bunescu, R.: The ohiot1dm dataset for blood glucose level prediction: Update 2020. CEUR workshop proceedings **2675**, 71 (2020)

[8] Prioleau, T., Bartolome, A., Comi, R., Stanger, C.: Diatrend: A dataset from advanced diabetes technology to enable development of novel analytic solutions. Scientific Data 2023 10:1 **10**, 1–9 (2023) https://doi.org/10.1038/s41597-023-02469-5

[9] Jaeb Center for Health Research: Type 1 Diabetes EXercise Initiative: The Effect of Exercise on Glycemic Control in Type 1 Diabetes Study. https://doi.org/10.25934/PR00008428