SAFETY-AWARE DEEP-RL FOR AUTOMATED INSULIN DELIVERY: TOWARD INCLUSIVE DIABETES CARE

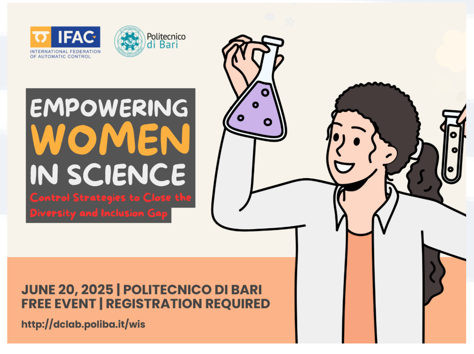
WORKSHOP EMPOWERING WOMEN IN SCIENCE: CONTROL STRATEGIES TO CLOSE THE DIVERSITY AND INCLUSION GAP

Giada Lops[1*], Gioacchino Manfredi[1†], Vito Andrea Racanelli[1‡], Luca De Cicco[1§], Saverio Mascolo[1¶]

[1]Politecnico di Bari, Bari, Italy

g.lops@phd.poliba.it, gioacchino.manfredi@poliba.it, luca.decicco@poliba.it, saverio.mascolo@poliba.it, vitoandrea.racanelli@poliba.it

0009-0005-7863-9183, 0000-0001-7644-0617, 0000-0002-8900-175X, 0000-0001-9686-5532, 0000-0002-9507-7398



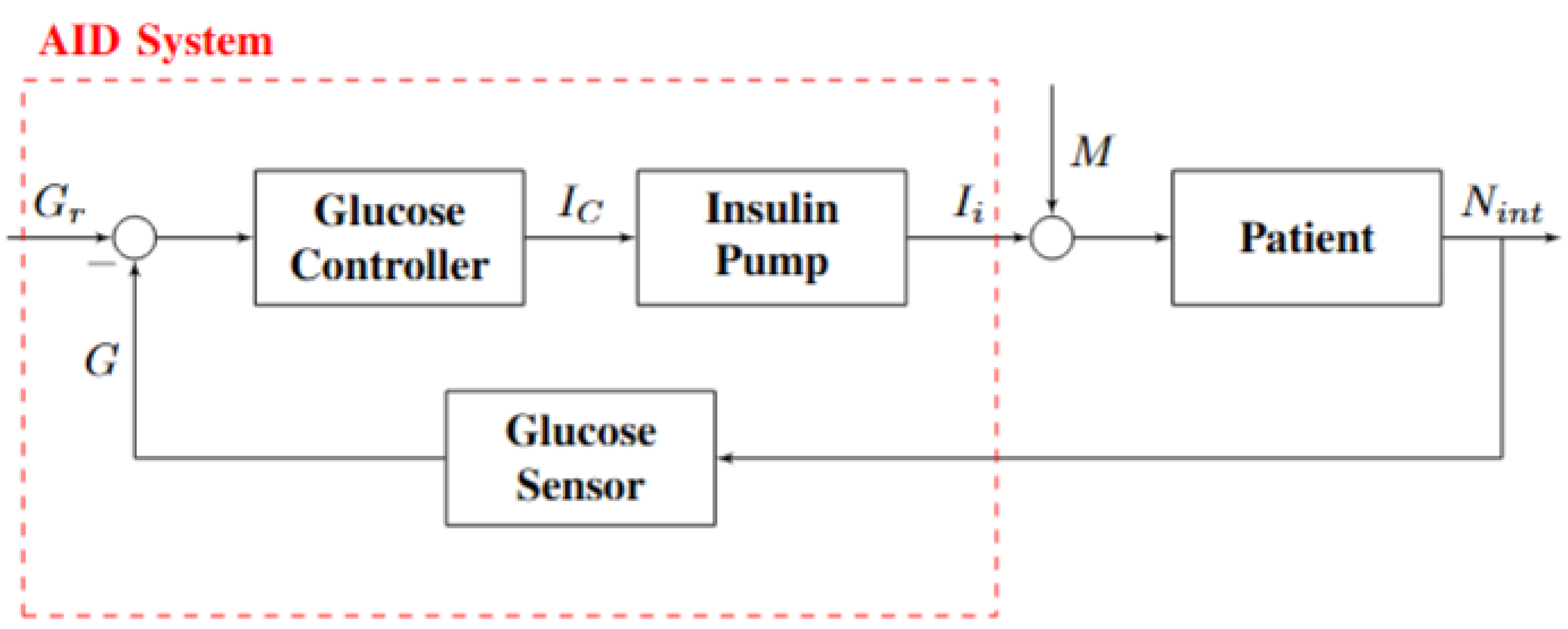
INTRODUCTION

Diabetes mellitus is a chronic disease characterized by elevated blood glucose, leading to long-term damage to vital organs. Type 2 diabetes results from insulin resistance or deficiency, while Type 1 is caused by autoimmune destruction of insulin-producing cells, requiring lifelong insulin therapy.

Maintaining glucose in a target range (70–180 mg/dL) is crucial. A key metric is *Time in Range* (TIR), with a goal of $\geq$70% per day. However, achieving this depends on several patient-specific factors and timely insulin delivery (Davidson *et al.*, 2008).

Automated Insulin Delivery (AID) systems integrate CGM sensors and insulin pumps with control algorithms like PID and MPC. Newer systems explore AI methods including Reinforcement Learning (RL) (Dénes-Fazakas *et al.*, 2024), Artificial Neural Networks, and predictive algorithms.

This work introduces a safe RL-based controller using *Maskable Proximal Policy Optimization* (Huang; Ontañón, 2020), which avoids unsafe actions by dynamically masking invalid insulin doses. Compared to standard PPO (Hettiarachchi *et al.*, 2024) and clinical practices, our method improves glucose control across diverse patient profiles, reducing risk and insulin use.



MASK FUNCTION DEFINITION

- **Action set:** $\mathcal{A} = \{a_1, \ldots, a_{n_T}\}$
- **Policy logits:** $\theta_i = l_i$
- **Validity conditions:**
  - If $G_t < 90$, then $a_1 = 0$ *(e.g., prevent insulin)*
  - If $G_t > 160$, then only the last n_v actions are valid
- **Mask Function:**

$$\mathrm{mask}(\theta_i) = \begin{cases} l_i, & \text{if } a_i \text{ is valid in } s \\ M, & \text{otherwise (e.g., } M = -\infty) \end{cases}$$

REWARD DESIGN BASED ON GLUCOSE LEVEL

- **Reward Function:**

$$r_t = \begin{cases} -3, & \text{if } G_t < 70 \\ +1, & \text{if } 70 \leq G_t \leq 180 \\ -1.5, & \text{if } G_t > 180 \end{cases}$$

MASKED PPO POLICY GRADIENT

Gradient with Masking:

$$g = \mathbb{E}_T \left[\sum_{t=0}^{T-1} \nabla_\theta \log \left(\pi'_\theta(a_t \mid s_t) \right) \sum_{k=0}^{\infty} \gamma^k r_{t+k} \right]$$

- The gradient update only uses **valid action logits** (according to the mask).
- This approach ensures **safety constraints** while maintaining **exploration**.

EVALUATION FRAMEWORK OVERVIEW

- **Simulator:** *Simglucose* (Xie, 2018), built on the FDA-approved *UVa/Padova* model
- **Devices:** *Dexcom CGM* and *Insulet insulin pump*
- **Target Glycemia:** 112 mg/dL *(minimizes risk index)*

RESULTS

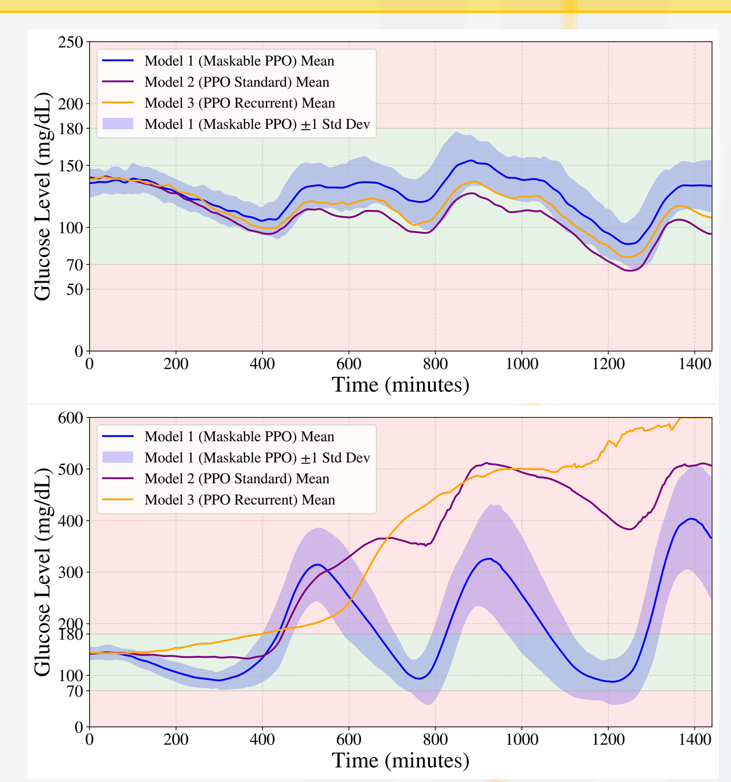
Evaluation Metrics for Maskable PPO, Standard PPO, and Recurrent PPO (Raffin *et al.*, 2021) (Pleines *et al.*, 2022) across Child, Adolescent, and Adult patients

Metric		Maskable PPO			Standard PPO			Recurrent PPO		
		Child	Adolescent	Adult	Child	Adolescent	Adult	Child	Adolescent	Adult
Time in Range [%]	Mean	54.43	96.85	96.39	38.52	87.42	92.65	37.70	68.25	95.60
	Median	54.17	98.33	98.33	33.41	87.60	92.71	36.21	68.85	96.46
	Min	40.00	84.79	76.25	25.78	71.46	78.96	19.10	35.00	83.96
	Max	71.25	100.00	100.00	67.20	99.17	100.00	68.90	96.04	100.00
Average Glucose Level [mg/dL]	Mean	189.06	115.05	126.09	297.62	95.69	108.28	290.40	85.58	115.39
	Median	189.91	115.17	125.70	304.14	95.11	108.54	291.50	85.33	115.76
	Min	157.36	106.51	115.42	210.69	87.11	90.85	223.77	74.66	96.49
	Max	214.54	125.65	135.01	381.95	107.06	124.84	344.97	95.75	135.63
Totale Injected Insulin [U]	Mean	4.35	13.52	17.32	2.79	16.81	21.25	1.52	18.95	19.90
	Median	4.36	13.58	17.27	3.11	16.83	21.21	1.56	18.97	19.85
	Min	2.48	12.01	15.96	1.62	15.81	19.91	1.07	17.95	18.56
	Max	4.83	15.02	19.16	3.47	17.77	22.64	2.00	19.84	21.26
Risk Index	Mean	15.89	1.47	1.55	35.17	3.05	1.97	32.83	6.12	1.38
	Median	16.16	1.42	1.38	35.97	3.02	1.73	33.11	5.91	1.31
	Min	6.91	0.55	0.63	17.09	1.33	0.62	17.81	3.10	0.53
	Max	23.03	3.67	3.56	52.41	5.33	5.24	45.37	10.68	4.62
Cumulative Rewards	Mean	-771.19	424.15	418.84	-1413.97	220.30	303.61	-734.47	-241.10	389.80
	Median	-772.00	450.00	456.75	-1636.00	232.25	332.00	-764.75	-216.50	412.00
	Min	-1262.00	159.50	111.00	-2082.50	-157.00	-227.00	-1317.00	-1088.00	17.00
	Max	-80.00	480.00	480.00	-340.00	464.00	480.00	-173.50	404.00	480.00

SIMULATION SETUP

- **CPU:** AMD Ryzen 5 7520U @ 2.8 GHz
- **Training:** ~496 s/patient/model; total ~1h 25m
- **Maskable PPO:** Most stable; reward ~120 (Child), ~100 (Adol.)
- **Recurrent PPO:** High peak in Adults (>400), unstable in others
- **Standard PPO:** Stable, low reward: ~70 (Adol./Adults), ~10 (Child)

BG TRACES



CONCLUSIONS

- Maskable PPO ensures valid insulin dosing, reducing risks and outperforming other PPO variants and clinical methods.
- Challenges remain for pediatric cases and when meal data is missing; emergency boluses may be needed.
- Future work: improve adaptability for diverse patients and scenarios.

REFERENCES

- DAVIDSON, Paul C. *et al.* Analysis of Guidelines for Basal-Bolus Insulin Dosing: Basal Insulin, Correction Factor, and Carbohydrate-To-Insulin Ratio. **Endocrine Practice**, 2008.
- DÉNES-FAZAKAS, Lehel *et al.* Reinforcement Learning: A Paradigm Shift in Personalized Blood Glucose Management for Diabetes. **Biomedicines**, 2024.
- HETTIARACHCHI, Chirath *et al.* G2P2C — A modular reinforcement learning algorithm for glucose control by glucose prediction and planning in Type 1 Diabetes. **Biomedical Signal Processing and Control**, 2024.
- HUANG, Shengyi; ONTAÑÓN, Santiago. A Closer Look at Invalid Action Masking in Policy Gradient Algorithms. **CoRR**, 2020.
- PLEINES, Marco *et al.* **Generalization, Mayhems and Limits in Recurrent Proximal Policy Optimization**. [*S. l.: s. n.*], 2022.
- RAFFIN, Antonin *et al.* Stable-Baselines3: Reliable Reinforcement Learning Implementations. **Journal of Machine Learning Research**, 2021.
- XIE, Jinyu. **Simglucose v0.2.1**. [*S. l.: s. n.*], 2018. Online. Disponível em: https://github.com/jxx123/simglucose.

ACKNOWLEDGEMENTS

SHIELD

DIPARTIMENTO DI INGEGNERIA ELETTRICA E DELL'INFORMAZIONE

SPEAKER

Giada Lops

Polytechnic University of Bari

g.lops@phd.poliba.it