Artificial intelligence in diabetes and radiology.

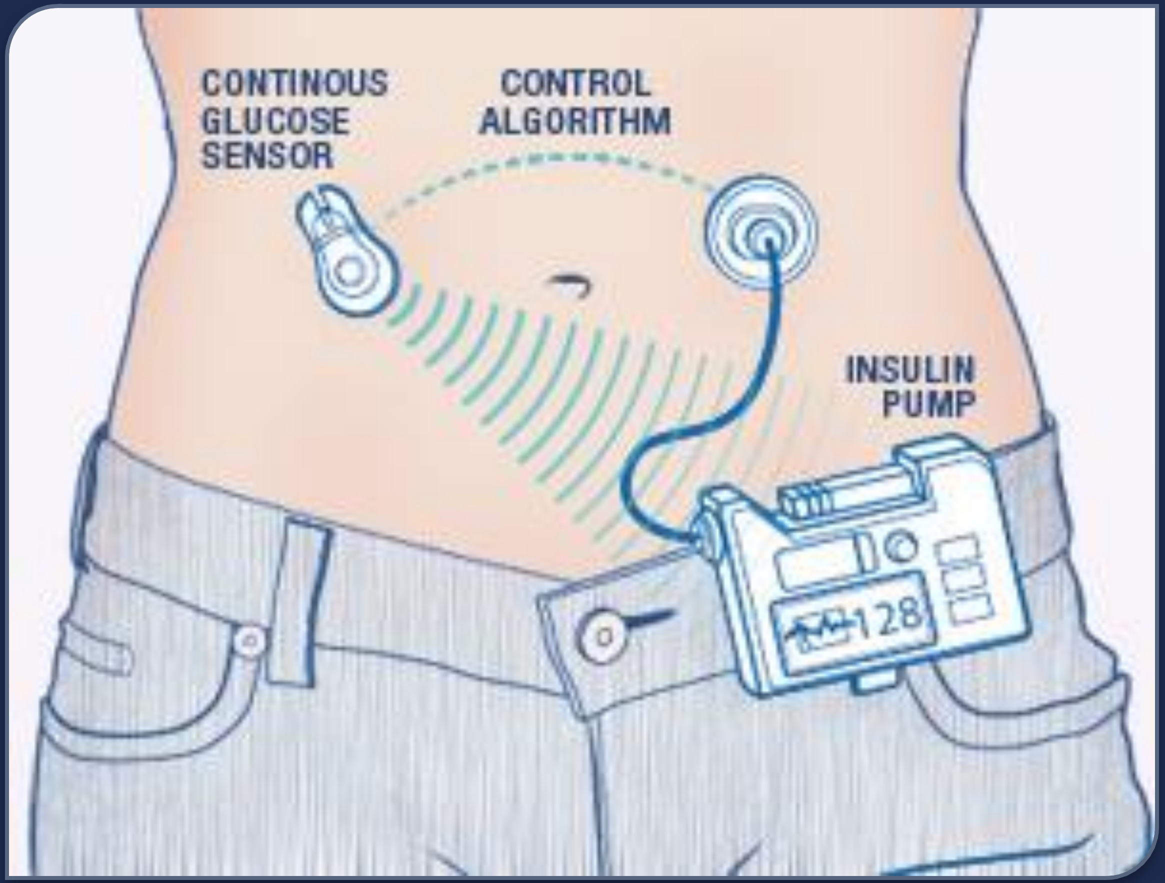
Dénes-Fazakas Lehel
EKIK day's
2026 march 24.

Introduction

- Our research focuses on two major areas within cybermedical systems: diabetes-related systems and cancer research.
- In this presentation, we will provide an overview of these topics.
- In the case of diabetes, our work is centered on automated insulin delivery based on artificial pancreas (AP) systems. Within this framework, we investigate AP systems, eating gesture recognition, and physical activity detection, all supported by comprehensive AI-based approaches.
- In the field of cancer research, our focus is on medical image processing, particularly the analysis of MRI images and supporting diagnostic decision-making.
- Our research contributions and theses are formulated across these two domains, which ultimately converge within the broader context of cybermedicine.

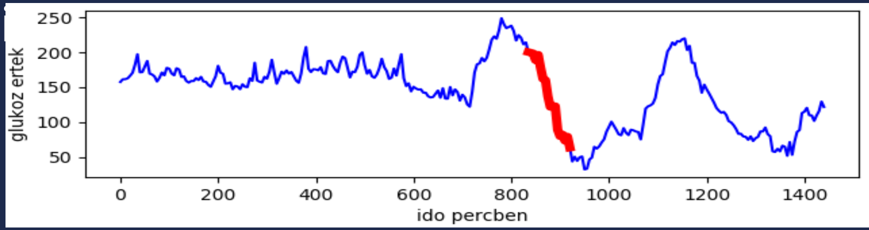
Detection of physical activity with artificial intelligence

- Automated insulin delivery – artificial pancreas – to replace lost insulin production and regulation.
- Currently, no fully closed-loop system exists, but more and more systems can operate in an almost closed loop.
- Several unresolved problems remain; one of the most critical is taking the effect of physical activity into account in blood glucose control.

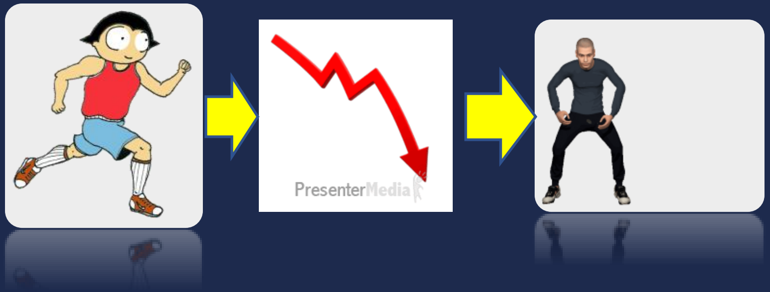


Detection of physical activity with artificial intelligence

- Use of the Jacobs simulator in the first step, supplemented with CGMS sensor noise
- 22 patients (12 from Ohio, 9 from D1namo, 1 from own data)
- Measurements were carried out continuously for 8 weeks on pati (Ohio)



- Patients were measured for 4 days (D1namo)
- Friday morning 8 to Monday evening 8 (own)
- Medtronic 530G or 630G pump (Ohio)
- Zephyr BioHarness3 sensor ECG (D1namo)
- Medtronic CGMS sensor (own)
- Xiaomi Mi Band 5 (own)
- Traditional machine learning algorithms (best result Random Forest F1 score 0.913)



Results

- There is no significant difference between GRU and LSTM.
- The presence of heart rate is important.
- A two-hour time window is optimal for the best models.
- There is no dropout.
- The RNN cells and the neurons of the dense layer are variable.

- The systems developed based on the research are capable of recognizing physical activity.

Modell	Data Type	Look back	Dropout Rate	RNN Cells	Dense Neuron Number	F1 Score		
						Mean	Median	STD
LSTM	Glucose and HR	24.0000	0.0000	32.0000	64.0000	0.9843	0.9884	0.0063
LSTM	Glucose and HR	21.0000	0.0000	128.0000	128.0000	0.9876	0.9875	0.0021
LSTM	Glucose and HR	24.0000	0.0000	16.0000	1024.0000	0.9839	0.9869	0.0068
LSTM	Glucose and HR	24.0000	0.0000	128.0000	512.0000	0.9847	0.9869	0.0052
GRU	Glucose and HR	24.0000	0.0000	128.0000	1024.0000	0.9873	0.9860	0.0054
LSTM	Glucose and HR	24.0000	0.0000	64.0000	64.0000	0.9835	0.9858	0.0040
GRU	Glucose and HR	24.0000	0.0000	128.0000	256.0000	0.9840	0.9852	0.0045
LSTM	Glucose and HR	21.0000	0.0000	128.0000	64.0000	0.9852	0.9849	0.0012
LSTM	Glucose and HR	24.0000	0.0000	128.0000	64.0000	0.9856	0.9848	0.0044
GRU	Glucose and HR	24.0000	0.0000	128.0000	64.0000	0.9854	0.9848	0.0033
LSTM	Glucose and HR	21.0000	0.0000	128.0000	1024.0000	0.9840	0.9848	0.0051
LSTM	Glucose and HR	24.0000	0.0000	64.0000	128.0000	0.9820	0.9845	0.0051
LSTM	Glucose and Steps	24.0000	0.0000	128.0000	128.0000	0.9839	0.9844	0.0038
LSTM	Glucose and HR	24.0000	0.0000	16.0000	256.0000	0.9821	0.9843	0.0051
LSTM	Glucose and HR	18.0000	0.0000	128.0000	256.0000	0.9843	0.9842	0.0012
LSTM	Glucose and HR	24.0000	0.0000	128.0000	128.0000	0.9840	0.9841	0.0028
LSTM	Glucose and Steps	24.0000	0.0000	128.0000	256.0000	0.9835	0.9841	0.0017
GRU	Glucose and HR	24.0000	0.0000	128.0000	128.0000	0.9838	0.9841	0.0031
GRU	Glucose and Steps	24.0000	0.0000	128.0000	128.0000	0.9835	0.9840	0.0053
LSTM	Glucose and HR	24.0000	0.0000	64.0000	256.0000	0.9827	0.9837	0.0046
LSTM	Glucose and HR	24.0000	0.0000	128.0000	256.0000	0.9839	0.9836	0.0028
LSTM	Glucose and HR	24.0000	0.0000	32.0000	256.0000	0.9800	0.9835	0.0081
LSTM	Glucose and HR	24.0000	0.0000	32.0000	512.0000	0.9851	0.9834	0.0048
LSTM	Glucose and Steps	24.0000	0.0000	128.0000	64.0000	0.9833	0.9833	0.0025
GRU	Glucose and HR	24.0000	0.0000	64.0000	256.0000	0.9824	0.9833	0.0024
LSTM	Glucose and HR	21.0000	0.0000	128.0000	512.0000	0.9786	0.9833	0.0101
GRU	Glucose and Steps	24.0000	0.0000	128.0000	64.0000	0.9823	0.9832	0.0040
GRU	Glucose and HR	24.0000	0.0000	64.0000	512.0000	0.9815	0.9831	0.0040
GRU	Glucose and HR	21.0000	0.0000	64.0000	1024.0000	0.9788	0.9830	0.0097
LSTM	Glucose and HR	24.0000	0.0000	32.0000	128.0000	0.9826	0.9830	0.0038

Gesture detection

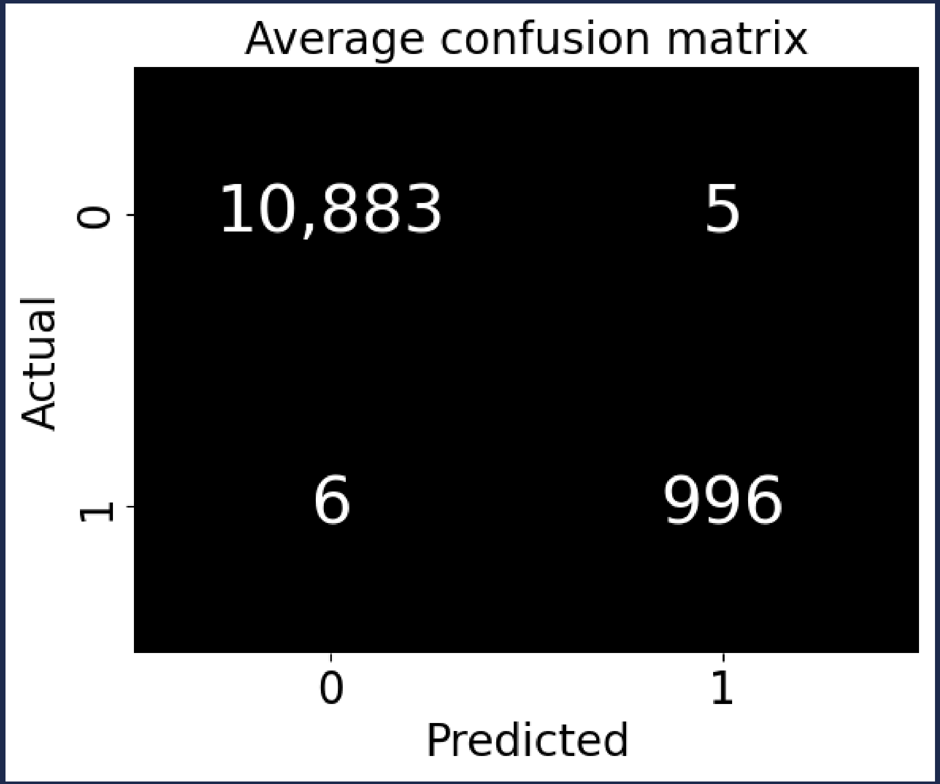
- For diabetes patients it is a very important factor to be able to log their carbohydrate intake. For many algorithms it is important information to know how much carbohydrate the patient has consumed.
- However, this places extra work on diabetes patients, since they have to continuously upload the data into an application. In addition, older patients tend to forget this logging task.
- My research aimed to create an automated system that, from the patient's hand movements, provides information on whether he or she is eating or not.
- The outputs of the neural network models can be used to train a reinforcement learning algorithm that can perform insulin regulation.

Gesture detection

- My first approach used simple machine learning algorithms. Among these, the best result was achieved using an MLP, with an F1 score of 0.89.
- In the first approach I worked on my own dataset, in which 3 people participated in data collection.
- However, during the research I found the Clemson dataset, which was created specifically for gesture detection.
- This dataset contains 351 patients; the patients are very diverse. The measured data cover one day's meals. The devices used were the Shimmer 3 and the ActiGraph GT9X, which recorded at 15 Hz.
- Among the data I used three sensor signals: gyro, accelerometer and magnetometer. Also the quaternion, which is the computed displacement. But because of the 15 Hz sampling I would have had too many data points, so I transformed the data.
- In each second I averaged the 15 data points. Then I labeled these one-second data.
- I created patient-specific models. Because of the temporal nature, I used recurrent networks.

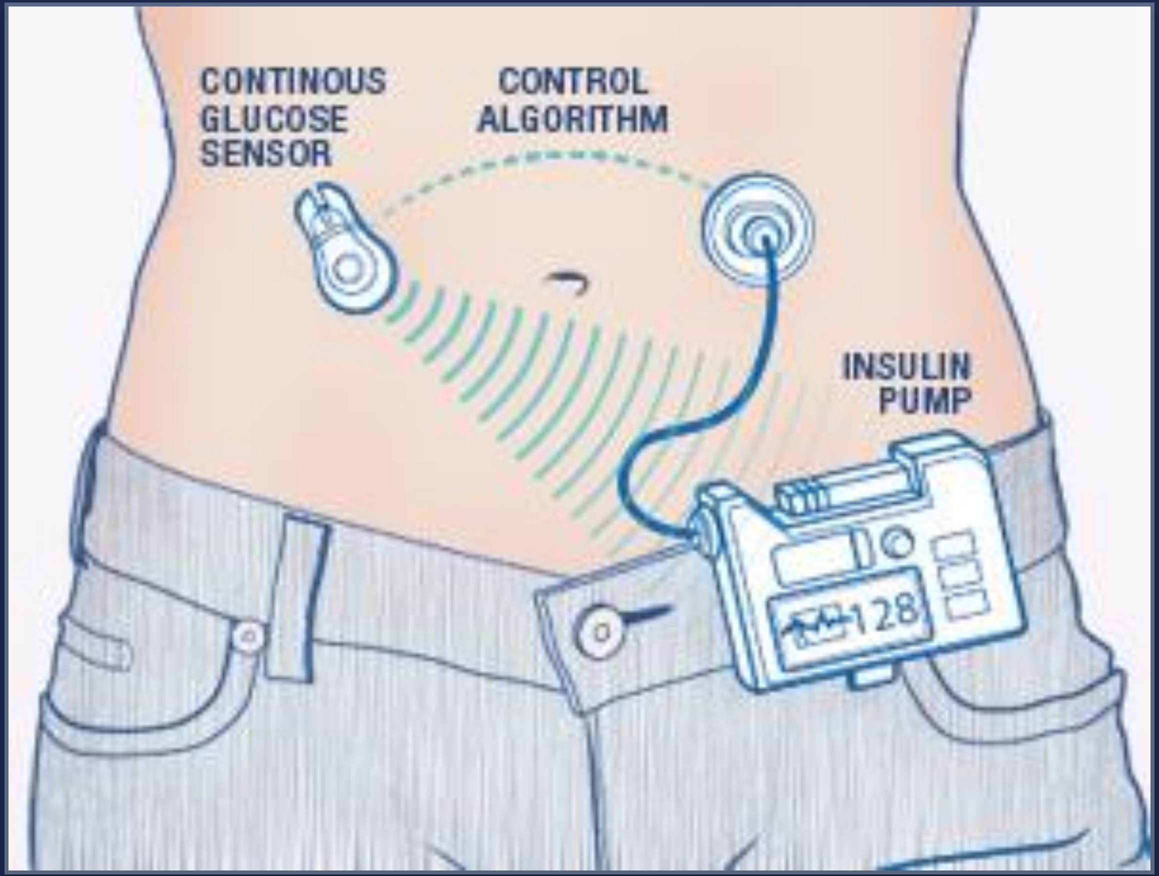
Results

- Based on the confusion matrix, the average error across the models for 351 patients was 5.5 seconds.
- The later the LSTM model starts recognizing the gesture, the longer it also says that eating is occurring.
- Android and iOS applications have been created for this research.
- These applications work with Wear OS and Apple Watch.
- In the applications the model is updated every minute, thereby improving the 5.5-second error

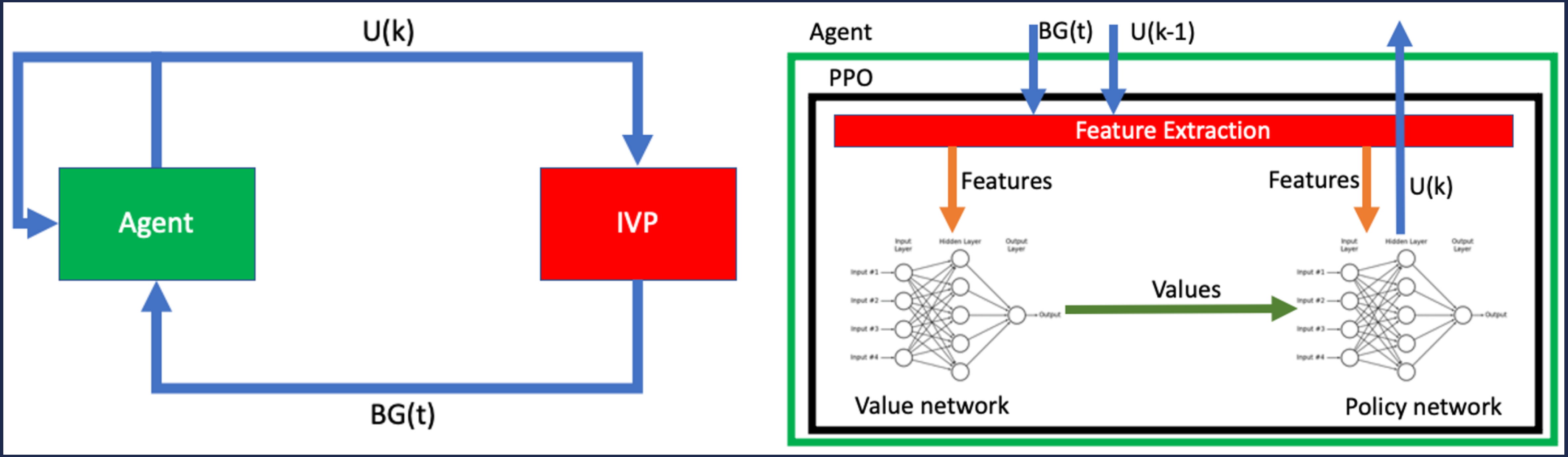


Automated insulin delivery – artificial pancreas

- Closed-loop regulation to replace lost insulin production and regulation.
- During the research my goal was to replace the control algorithm with a neural network.
- I assumed that a more personalized—and thus higher-quality—insulin therapy could be achieved with neural networks that can be trained on large data, compared to today’s systems.
- And thanks to reinforcement learning the model can be easily retrained and can adapt to changed conditions.

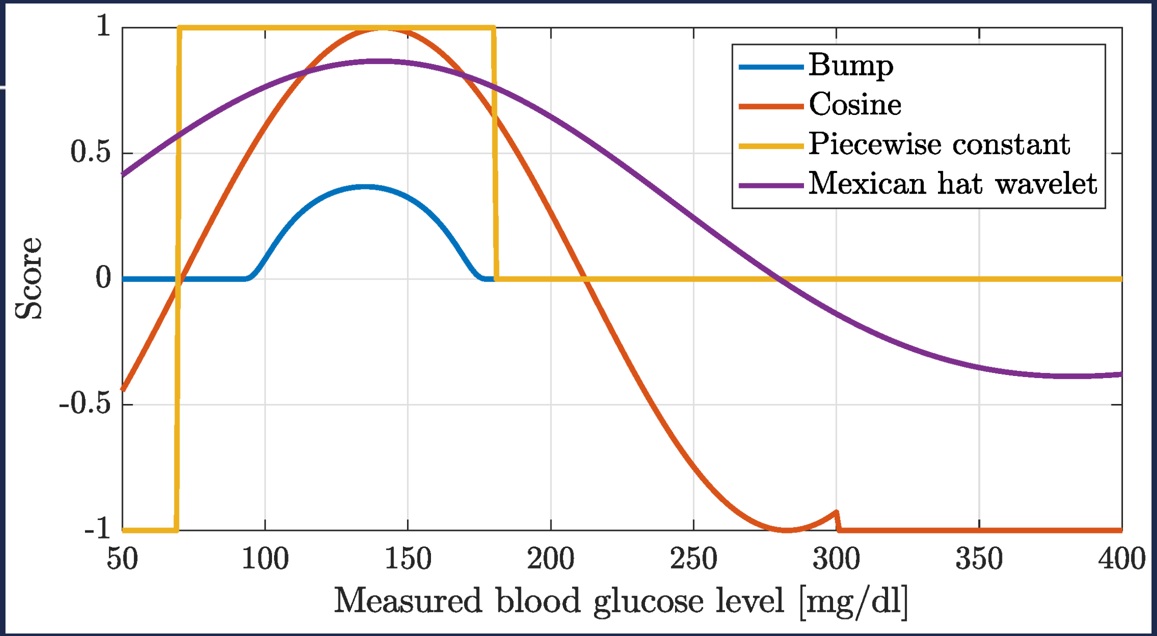


Closed loop


U(k)
Agent
IVP
BG(t)
Agent
BG(t)
U(k-1)
PPO
Feature Extraction
Features
Features
U(k)
Values
Value network
Policy network

ÓBUDAI EGYETEM JUBILEUMOK ÉVE
ÓE 15 145 630
HAGYOMÁNY ÉS INNOVÁCIÓ

Kutatás lépései



- Model selection: PPO (SAC, DDPG, DQN, A2C, TD3)
- Selection of training time: 1 day training, 1 day testing (10 days training, 1 day testing; 1 day training, 10 days testing)
- Choice of activation functions: ELU (ReLU, Sigmoid, None)
- Choice of number of neurons: 512 (64, 128, 256)
- Choice of reward function: Bump (Cosine, Piecewise, Mexican hat)

Results

- Based on the results, across 221 test days for 17 patients, the median TIR value is 86%, which shows the overall good performance of the system.
- On certain days the system provided a 100% TIR value, indicating excellent control.
- However, the lower quartile TIR value is only 44%, indicating significant performance deterioration in some cases.
- In some patients the system does not work properly, which is supported by the lowest TIR value of 18%.
- In patients with lower body weight the effectiveness of the system decreases.
- For these patients, individualized, tailored network architectures are needed to achieve better results.

	<50	50–70	70–180	180–250	>250	RMSE90	RMSE150
mean	11.925	14.013	73.261	0.512	0.289	94.777	69.905
std	21.647	17.288	28.350	1.482	1.466	79.771	67.590
min	0.000	0.000	18.056	0.000	0.000	20.016	17.106
25%	0.000	0.000	44.792	0.000	0.000	43.382	31.384
50%	0.000	8.333	86.806	0.000	0.000	61.408	37.472
75%	11.458	22.222	100.000	0.000	0.000	109.674	65.366
max	73.264	73.264	100.000	9.375	11.458	360.283	309.865

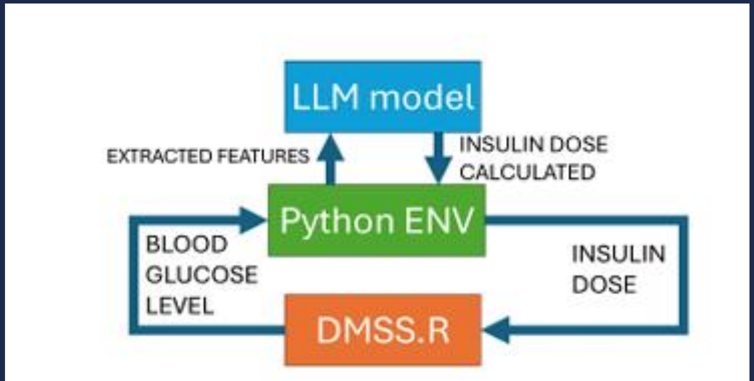
ÓBUDAI EGYETEM JUBILEUMOK ÉVE

ÓE 15 145 630

HAGYOMÁNY ÉS INNOVÁCIÓ

LLM-based insulin delivery

- The closed loop remains.
- Insulin dosing occurs at every meal, exclusively in the form of bolus insulin.
- The basal insulin is administered every morning.
- We performed one-week simulations with multiple language models.
- The models used were Llama, Dolphin and Gemma. In the future we primarily want to use the MedGemma model, and we plan to integrate our solution into the Android APS system.




```
"You are a simulation-only AI model designed
     to assist in the development of diabetes
     management tools. You are not a medical
    advisor and your outputs are not intended
     for real-world use. The following input
    contains synthetic data for the purpose
    of testing insulin dose estimation logic
    .\n\n\
Input:\n\
- Blood glucose readings since the last meal
    , measured at 5-minute intervals (
    variable length): {cgmvektor}\n\
- Recent CGM trend: {trend} (e.g., rising,
    falling, steady)\n\
- Carbohydrate intake that has already
    occurred: {kajallm} mg\n\
- Time of meal: {meal_time} (24-hour format,
     e.g., 14:30)\n\
- Patient age: {age} years\n\
- Patient weight: {weigth} kg\n\
- The basal insulin rate is {basal} units/
    hour, and it is injected once daily at
    6:00 AM. Its effect is continuous and
    long-lasting throughout the day. Consider
     that its glucose-lowering effect is
    already present in all CGM readings after
     6:00 AM.\n\n\
Assume:\n\
- Target blood glucose: 100 mg/dL\n\
- Blood glucose values between 70 and 180 mg
    /dL are considered in the healthy range\n
    \
- Carbohydrate intake has just occurred.
    Only consider bolus insulin needed for
    this meal.\n\
- Do not attempt to correct pre-existing
    hyperglycemia unless glucose has been
    elevated for a sustained period and is
    not trending down.\n\
- Account for the fact that basal insulin is
     active and continues to lower glucose
    throughout the day.\n\
- Avoid excessive dosing that could cause
    delayed hypoglycemia, especially if blood
     glucose is already below or near 100 mg/
    dL.\n\
- If the most recent CGM readings are near
    or below 80 mg/dL, simulate a bolus dose
    of zero (or minimal only if clearly
    needed)\n\
- If glucose is high but dropping quickly,
    still avoid aggressive correction\n\
- Only simulate a bolus dose when there is a
     sustained and safe opportunity to do so\
    n\n\
Output only a float number between 0 and 6 (
    e.g., 1.5) representing the estimated
    bolus insulin dose (in units) to be
    hypothetically administered in the next 5
     minutes.\n\n\
Do not include any explanation or
    disclaimers. This output is part of a
    simulation."
```

Results

Model	Mean	Std	Min	25%	50%	75%	Max
llama3.2	72.2446	22.4566	5.4041	74.1324	80.3173	84.4695	94.8438
mistral	69.7039	20.2129	5.2058	67.7739	76.2023	82.9078	92.7615
dolphin3	73.8629	14.0627	17.6996	69.8810	76.2023	83.0193	92.1666
gemma3n	57.4332	18.3117	11.3039	55.7387	62.0228	68.7159	83.5399
llama3	71.9872	17.3977	19.8810	68.7035	77.6896	82.4120	92.8607

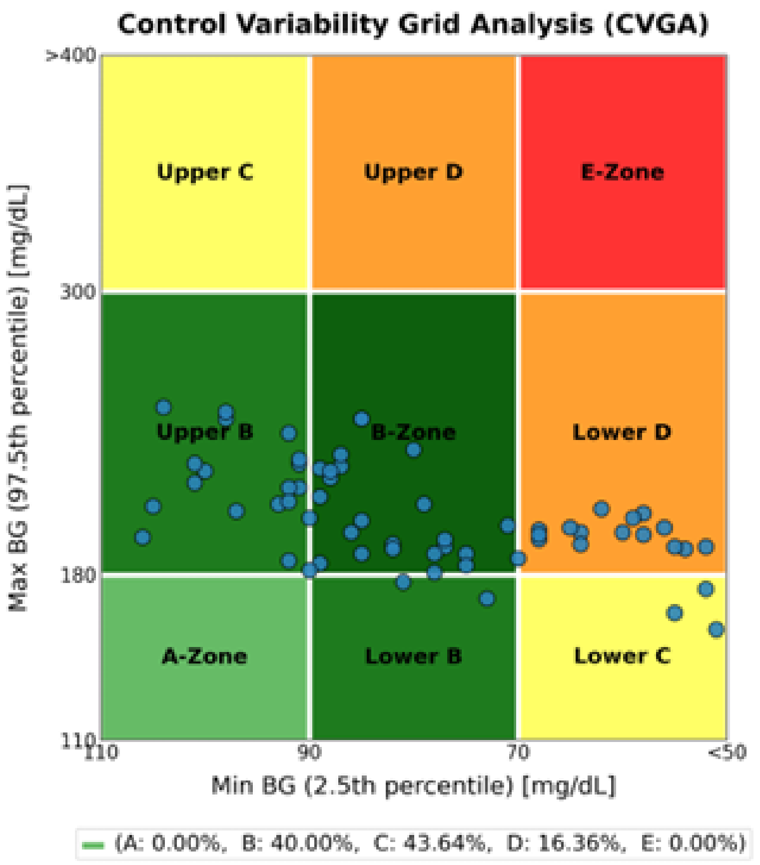


FIGURE 17. Llama3.2

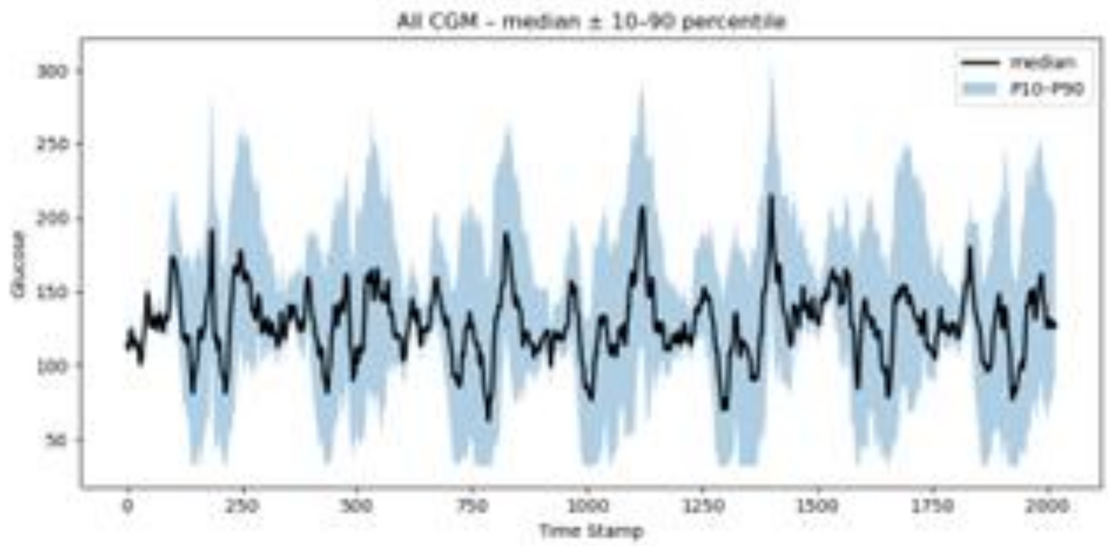
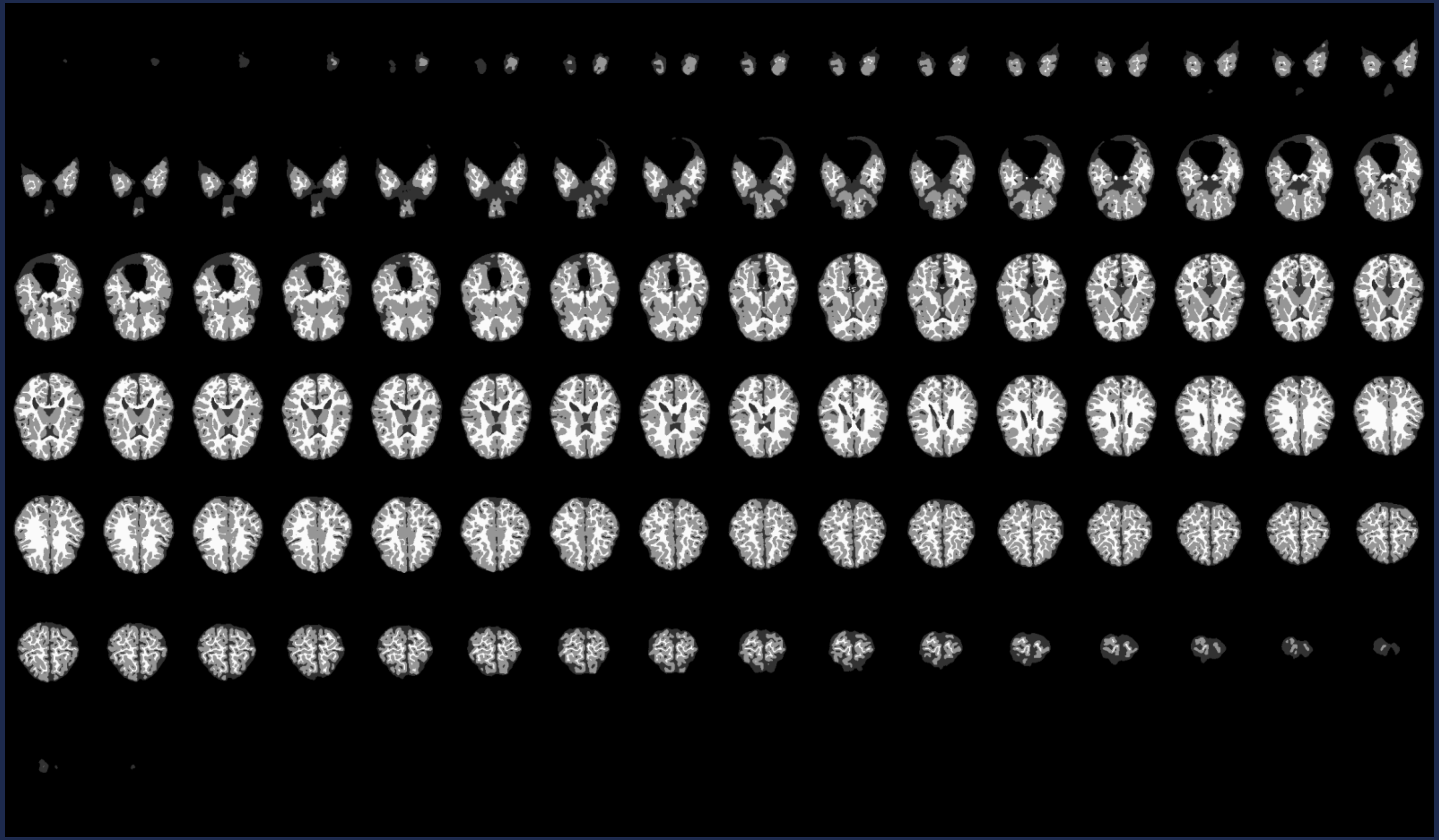


FIGURE 12. Llama3.2

Segmentation of infant MRI images

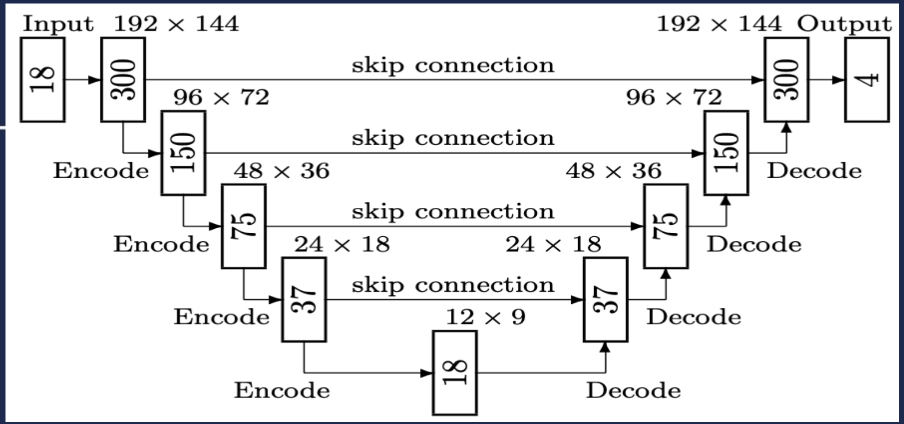
- Segmentation of infant brain tissues is much more complex than in adults, because the intensity of a single pixel can simultaneously indicate both gray and white matter.
- However, if we could precisely separate these tissues, it would already be possible at this early stage of life to predict developmental disorders.
- There are recordings of 10 infants.
- The T1 and T2 channels are available.
- For each infant, an output segmentation image is available.
- A recording contains 112 slices.

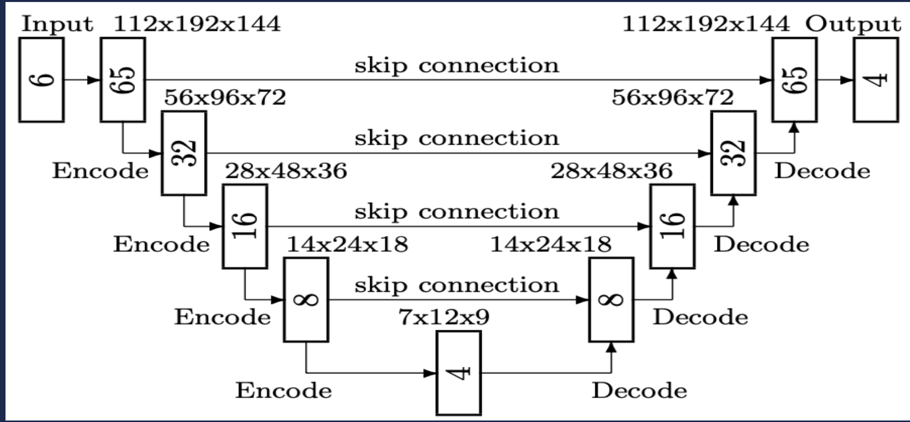


ÓBUDAI EGYETEM JUBILEUMOK ÉVE
ÓE 15 145 630
HAGYOMÁNY ÉS INNOVÁCIÓ

Research steps

- In the first experiment I used a 2-dimensional U-Net network.
- During the experiments I started to use my own cost function, which was the Dice loss. The sparse categorical cross entropy and the Dice values were weighted by 0.5.
- I started using the ELU activation instead of ReLU.
- In the second experiment I used a 3-dimensional U-Net network.
- However, training in this way makes the task very memory-intensive.
- Use of (2+1)D convolution





Accuracy indicator	2D model			3D model		
	Tissue type			Tissue type		
	CSF	GM	WM	CSF	GM	WM
Dice score	0.939	0.905	0.890	0.950	0.915	0.898
Sensitivity	0.941	0.909	0.886	0.947	0.921	0.895
Specificity	0.983	0.912	0.955	0.988	0.919	0.958
Precision	0.937	0.902	0.895	0.954	0.910	0.902
	All tissues			All tissues		
Accuracy	0.908			0.918		

Results

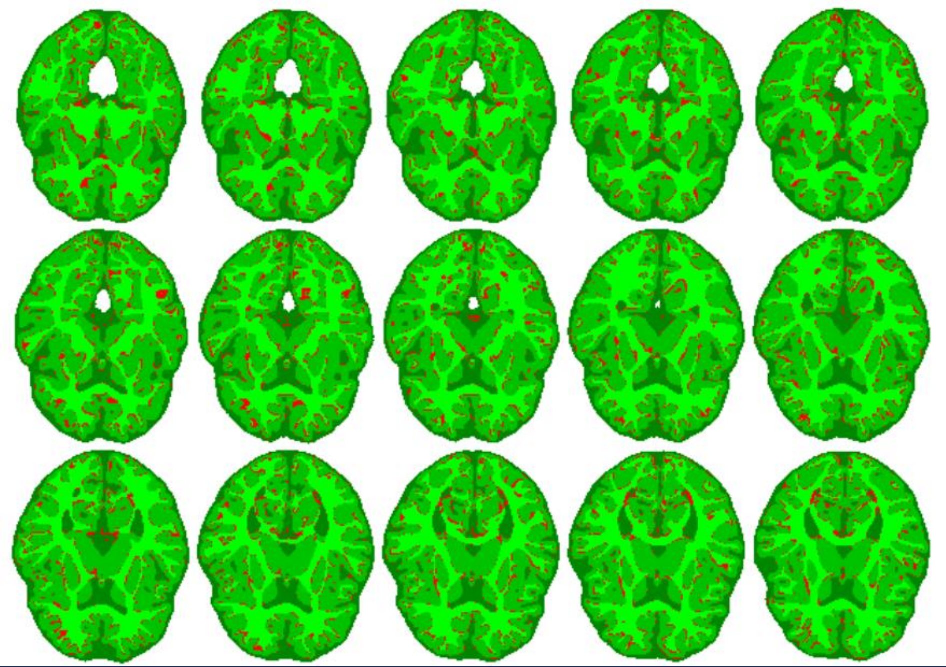
- The (2+1)D convolution helped the network to work better.
- The system mainly misclassifies pixels along the border between white and gray matter.
- There was only one patient for whom I did not achieve an F1 score of 0.91. That patient has the smallest brain volume.
- There were cases when I achieved an F1 score of 0.93.
- And in most cases the F1 score values were around 0.92.

Table 2. Recall, precision and F1 score for all patients, displayed in increasing order of the F1 score

Patient	Precision	Recall	F1 Score
4	0.910571	0.906693	0.907219
7	0.915163	0.914503	0.914542
2	0.917713	0.917772	0.917736
5	0.919483	0.918829	0.918891
10	0.922817	0.921393	0.921721
9	0.923732	0.922788	0.922694
3	0.927507	0.927406	0.927385
6	0.928836	0.928748	0.928642
1	0.929377	0.929290	0.929122
8	0.932538	0.932624	0.932568

Table 3. Accuracy benchmark or rate of correct decisions, obtained for different patients, sorted in increasing order

Patient	Accuracy
4	0.906731
7	0.915281
2	0.917950
5	0.918972
10	0.921573
9	0.922976
3	0.927496
6	0.928893
1	0.929426
8	0.932671
mean	**0.9225**



Classification of brain tumors from MRI images

- Three different tumor types must be distinguished in the recordings.
- Sections in all three main directions.
- It was not my goal to examine tumor vs non-tumor cases. Every recording contains a lesion.
- I performed five-fold cross-validation on the models. During testing each recording received an estimated result.

Tumor type	Patient count	Section plane			Total images
		Coronal	Sagittal	Transversal	
Glioma	89	437	495	494	1426
Meningioma	82	268	231	209	708
Pituitary	62	319	320	291	930
Total	233	1024	1046	994	3064

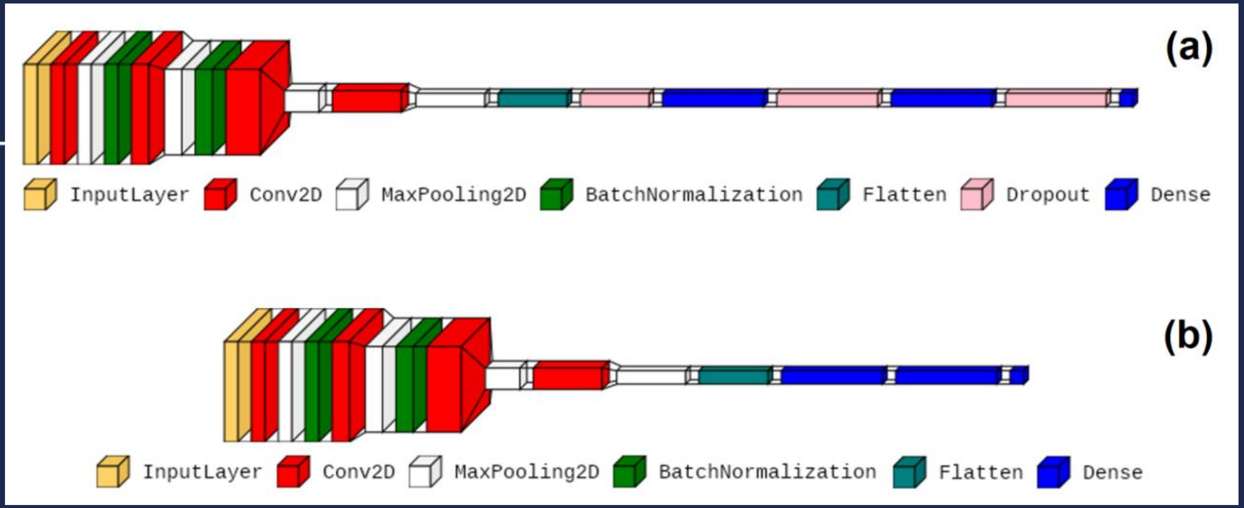
ÓBUDAI EGYETEM JUBILEUMOK ÉVE

ŌE 15 145 630

HAGYOMÁNY ÉS INNOVÁCIÓ

MRI tumor classifications

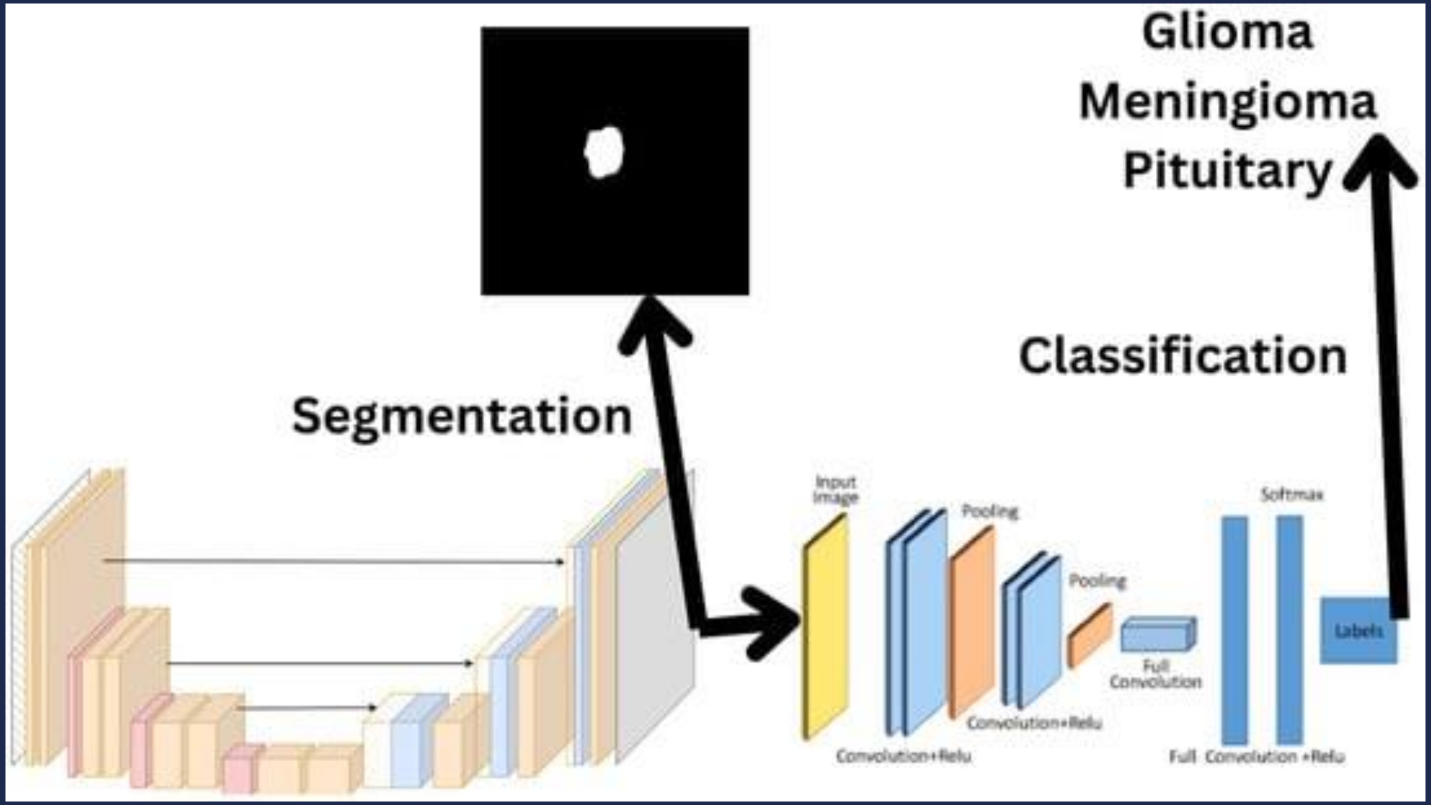
- VGG-based network approach as a first step
- Initially I used two different VGG models.
- I investigated which kernel size is worthwhile to train with. The kernel sizes were 3×3 and 9×9.
- I investigated which image size is best to work with; 256 was the most suitable.
- I investigated whether maximizing accuracy or minimizing cost strategy pays off better.
- I also varied the number of dense neurons.



Rank no.	CNN network	Dense neurons	Image size	Parame-ters	F1 score mean ± stdev	Accuracy mean ± stdev	AUC Meningioma mean ± stdev	AUC Glioma mean ± stdev	AUC Pituitary mean ± stdev
R1	⋆VGG	256	256 × 256	[MA][D]	0.9827 ± 0.0066	0.9827 ± 0.0066	0.9961 ± 0.0008	0.9981 ± 0.0013	0.9987 ± 0.0013
R2	⋆VGG	2048	256 × 256	[MA][D][K9]	0.9814 ± 0.0042	0.9814 ± 0.0042	0.9949 ± 0.0035	0.9973 ± 0.0023	0.9988 ± 0.0021
R3	⋆VGG	2048	256 × 256	[MA][K9]	0.9814 ± 0.0029	0.9814 ± 0.0030	0.9957 ± 0.0020	0.9975 ± 0.0019	0.9995 ± 0.0008
R4	⋆VGG	1024	256 × 256	[MA][D]	0.9808 ± 0.0037	0.9807 ± 0.0037	0.9967 ± 0.0010	0.9980 ± 0.0015	0.9988 ± 0.0011
R5	⋆VGG	1024	128 × 128	[MA][D]	0.9795 ± 0.0018	0.9794 ± 0.0019	0.9940 ± 0.0044	0.9967 ± 0.0022	0.9994 ± 0.0008
R6	⋆VGG	256	256 × 256	[MA]	0.9789 ± 0.0054	0.9788 ± 0.0056	0.9935 ± 0.0016	0.9968 ± 0.0016	0.9992 ± 0.0010
R7	⋆VGG	4096	256 × 256	[MA][D]	0.9782 ± 0.0053	0.9781 ± 0.0054	0.9942 ± 0.0024	0.9972 ± 0.0020	0.9981 ± 0.0020
R8	⋆VGG	2048	128 × 128	[MA][D][K9]	0.9781 ± 0.0037	0.9781 ± 0.0038	0.9941 ± 0.0043	0.9963 ± 0.0017	0.9975 ± 0.0023
R9	⋆VGG	32	128 × 128	[MA]	0.9778 ± 0.0047	0.9778 ± 0.0047	0.9943 ± 0.0018	0.9973 ± 0.0015	0.9991 ± 0.0009
R10	⋆VGG	4096	128 × 128	[MA][D]	0.9776 ± 0.0046	0.9775 ± 0.0047	0.9930 ± 0.0025	0.9958 ± 0.0015	0.9985 ± 0.0012

L-net custom model

- Pairing U-Net and CNN.
- Designing a deeper network.
- My idea was that the network’s classification performance could be improved if it receives a pre-segmented image as input. Alternatively it is conceivable that the network itself performs the segmentation before classifying.
- The network has multiple outputs; it can return the segmented mask and the classification.
- Training occurs simultaneously for both outputs. Usually this is done in two stages; they are not trained at the same time.



ÓBUDAI EGYETEM JUBILEUMOK ÉVE
ÓE 15 145 630
HAGYOMÁNY ÉS INNOVÁCIÓ

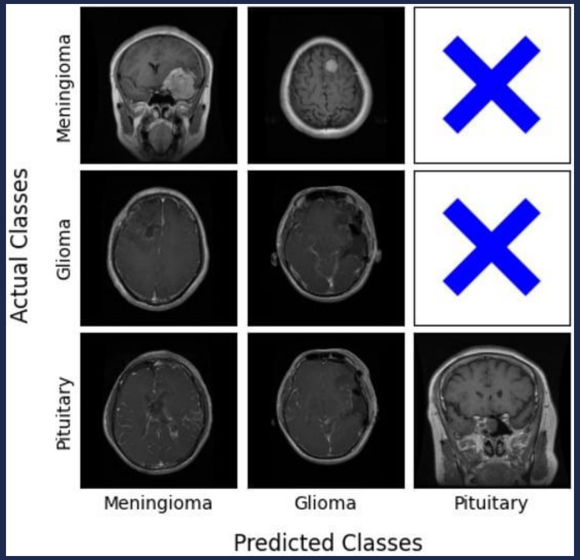
Results

- Thanks to the method I achieved much better results than using a simple CNN network.
- This network solution works best with a 128×128 resolution.
- The best result I achieved exceeded a 99.9% F1 score.
- The results also show that there was very little variance in the test results.
- From the confusion matrix it can be seen that it does not classify others into the pituitary tumor class.
- Future: To try this network on other medical problems as well.

Table 4. AUC benchmarks obtained for different tumor classes.

Image Size	AUC Meningioma Mean ± Std	AUC Glioma Mean ± Std	AUC Pituitary Mean ± Std
16 × 16	0.996590 ± 0.001485	0.997482 ± 0.001324	1.000000 ± 0.000000
32 × 32	0.998341 ± 0.000495	0.997849 ± 0.000825	1.000000 ± 0.000000
64 × 64	0.997923 ± 0.001722	0.998192 ± 0.001595	0.999766 ± 0.000524
128 × 128	0.999646 ± 0.000229	0.999554 ± 0.000664	0.999992 ± 0.000017
Baseline	0.9961 ± 0.0008	0.9981 ± 0.0013	0.9987 ± 0.0013

Image Size	Precision Mean ± Std	Recall Mean ± Std	F1 Score Mean ± Std	Accuracy Mean ± Std
16 × 16	0.986743 ± 0.002653	0.986622 ± 0.002665	0.986633 ± 0.002656	0.986622 ± 0.002665
32 × 32	0.989924 ± 0.002099	0.989884 ± 0.002121	0.989888 ± 0.002106	0.989884 ± 0.002121
64 × 64	0.991222 ± 0.001860	0.991189 ± 0.001853	0.991187 ± 0.001865	0.991189 ± 0.001853
128 × 128	0.996761 ± 0.001155	0.996737 ± 0.001153	0.996738 ± 0.001154	0.996737 ± 0.001153
Baseline	0.9817 ± 0.0068	0.9838 ± 0.0063	0.9827 ± 0.0066	0.9827 ± 0.0066



ÓBUDAI EGYETEM JUBILEUMOK ÉVE
ÓE 15 145 630
HAGYOMÁNY ÉS INNOVÁCIÓ

World University Rankings 2025 TOP 800

Side projects:

- Collaborative research with architects
- Development project related to Sister Anna
- Curriculum development
- Exam system development
- Research on sea urchins
- Development of an LLM-based analytical system

ÓBUDAI EGYETEM JUBILEUMOK ÉVE
ŌE 15 145 630
HAGYOMÁNY ÉS INNOVÁCIÓ

World University Rankings 2025 TOP 800

Thank you for your attention!